ABSTRACT

Title of dissertation: DYNAMICS OF LARGE SYSTEMS OF NONLINEARLY EVOLVING UNITS

Zhixin Lu, Doctor of Philosophy, 2017

Dissertation directed by: Professor Edward Ott
Department of Physics;

The dynamics of large systems of many nonlinearly evolving units is a general research area that has great importance for many areas in science and technology, including biology, computation by artificial neural networks, statistical mechanics, flocking in animal groups, the dynamics of coupled neurons in the brain, and many others. While universal principles and techniques are largely lacking in this broad area of research, there is still one particular phenomenon that seems to be broadly applicable. In particular, this is the idea of emergence, by which is meant macroscopic behaviors that "emerge" from a large system of many "smaller or simpler entities such that ... large entities" [i.e., macroscopic behaviors] arise which "exhibit properties the smaller/simpler entities do not exhibit." [1]. In this thesis we investigate mechanisms and manifestations of emergence in four dynamical systems

consisting many nonlinearly evolving units. These four systems are as follows.

(a) We first study the motion of a large ensemble of many noninteracting particles in a slowly changing Hamiltonian system that undergoes a separatrix crossing. In such systems, we find that separatrix-crossing induces a counterintuitive effect. Specifically, numerical simulation of two sets of densely sprinkled initial conditions on two energy curves appears to suggest that the two energy curves, one originally enclosing the other, seemingly interchange their positions. This, however, is topologically forbidden. We resolve this paradox by introducing a numerical simulation method we call "robust" and study its consequences.

(b) We next study the collective dynamics of oscillatory pacemaker neurons in Suprachiasmatic Nucleus (SCN), which, through synchrony, govern the circadian rhythm of mammals. We start from a high-dimensional description of the many coupled oscillatory neuronal units within the SCN. This description is based on a forced Kuramoto model. We then reduce the system dimensionality by using the Ott-Antonsen Ansatz and obtain a low-dimensional macroscopic description. Using this reduced macroscopic system, we explain the east-west asymmetry of jet-lag recovery and discus the consequences of our findings.

(c) Thirdly, we study neuron firing in integrate-and-fire neural networks. We build a discrete-state/discrete-time model with both excitatory and inhibitory neurons and find a phase transition between avalanching dynamics and ceaseless firing dynamics. Power-law firing avalanche size/duration distributions are observed at critical parameter values. Furthermore, in this critical regime we find the same power-law exponents as those observed from experiments and previous, more re-

stricted, simulation studies. We also employ a mean-field method and show that inhibitory neurons in this system promote robustness of the criticality (i.e., an enhanced range of system parameter where power-law avalanche statistics applies).

(d) Lastly, we study the dynamics of "reservoir computing networks" (RCN's), which is a recurrent neural network (RNN) scheme for machine learning. The advantage of RCN's over traditional RNN's is that the training is done only on the output layer, usually via a simple least-square method. We show that RCN's are very effective for inferring unmeasured state variables of dynamical systems whose system state is only partially measured. Using the examples of the Lorenz system and the Rössler system we demonstrate the potential of an RCN to perform as an universal model-free "observer".

DYNAMICS OF LARGE SYSTEMS OF NONLINEARLY EVOLVING UNITS

by

Zhixin Lu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Professor Edward Ott, Chair
Professor Thomas M. Antonsen
Professor Michelle Girvan
Professor Christopher Jarzynski
Professor Rajarshi Roy

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor and mentor Prof. Edward Ott for his continuous support during my Ph.D study and research, for his patience, motivation, and immense knowledge. He was my dream advisor when I applied for grad school and will always be my role model. I feel lucky to have had the privilege to study and to do research with him. I have learned so much from him and we have shared so many memories in these years. Without him I could not have explored these interesting questions in the field of nonlinear dynamics. Those memories of our sitting together talking about research, editing our manuscripts, as well as chatting will forever be with me. I thank him for guiding me through my PhD years with so much care and support.

I would also like to thank the rest of my thesis committee: Prof. Michelle Girvan, Prof. Christopher Jarzynski, Prof. Thomas M. Antonsen, and Prof. Rajarshi Roy, without whom I could not have gotten this far. Specifically, I would like to thank my co-advisor Prof. Michelle Girvan for her enormous encouragement and guidance during my Ph.D study. The path to my defense was made much easier with her kind advice and encouragement. I have learned from her not only methods and theories of complex networks but also skills to communicate effectively in academia. I would also like to thank Prof. Christopher Jarzynski, who co-mentored me on my very first project in Maryland. Thank you for providing me the opportunity to join your research group so that I could start my very first project in my Ph.D study. Without your precious support I might not have been able to start my nonlinear

dynamics research and may eventually had a totally different career path. Chris is also the mentor to my twin brother. Both of us have learned so much from him. I would also like to express my sincere gratitude to Prof. Thomas M. Antonsen and Prof. Rajarshi Roy for their insightful comments and encouragement.

Thank you Prof. Brian Hunt for your help and inspiring discussions through which I learned so much in my reservoir computing project. You are the first mathematician I worked with, and I enjoyed all of our discussions so much. I can not thank you enough for your kind support and help. I am so happy to have had the chance to share with you the joy and surprise as we explored the amazing dynamics of reservoir computing networks. My sincere thanks also goes to Prof. John Weeks, Prof. Millard H. Alexander, Prof. James A. Yorke, Prof. Dave Levermore, Prof. Victor Yakovenko, Prof. Garegin Papoian, and Prof. Daniel Lathrop, with whom I have had valuable and joyful disscussions.

Special thanks go to Prof. Michael A. Coplan and Debbie Jenkins. You made me feel like I was in a big warm family in the chemical physics program in IPST. Your kind and caring help and support made my Ph.D study such a sweet and unforgetable memory that I will treasure forever. I would also like to thank my colleagues and good friends, Shane Squires, Wai Lim Ku, Matt Herrera, David Winogradoff, Matt Murray, Jaideep Pathak, Sarthak Subhankar, Nightvid Cole, Yi-Hsieh Wang, Yang Shen, Hao Wu, Haiqing Zhao, Huashuai Qu, Xuan Liu, Ke Ma, Jinxin Xue, Marie Chou, Jong Jun Lee, Kara Y.D., Xuan Yao, Chen Li, Hailong Bai, Jianbing Zhang, Zhe Wen, Zhe Wu, and many others. I feel so blessed to have had met you. You have made my life in UMD so bright and so joyful.

Special thanks go to my twin brother Zhiyue Lu who gave me so much company in Maryland. Countless happiness and joy from you and your wife Huizhi made those days in UMD so colorful and unforgettable. Thank you Mom and Dad. Thank you for trusting me all the way with so much love and support. I would also like to thank my wife Ying Wang, for her enormous love and the trust she places in me. Thank you for sharing your life with me. You taught me how to love and made my life complete. I also want to thank all my friends, brothers and sisters in the UMBC Bible study group and in the Chinese Bible Church of Baltimore County. I thank you all for sharing such great fellowship with me in the Lord.

Thank you my dear LORD for being my light and my salvation. Through Jesus Christ I found the peace that the world cannot give and the peace that the world cannot take away.

Table of Contents

List of Figures

List of Abbreviations

SCN	Suprachiasmatic nucleus,
RCN	Reservoir computing network,
RNN	Recurrent neural network,
RMS	Root-mean-square,
IREAP	Institute for Research in Electronics and Applied Physics,
IPST	Institute for Physical Science and Technology,

Chapter 1: Introduction

1.1 Overview

The Dynamics of large systems of many nonlinearly evolving units is a general research area that has great importance for many areas in science and technology, including biology, computation by artificial neural networks, statistical mechanics, flocking in animal groups, the dynamics of coupled neurons in the brain, and many others. While universal principles and techniques are largely lacking in this broad area of research, there is still one particular phenomenon that seems to be broadly applicable. In particular, this is the idea of emergence, by which is meant macroscopic behaviors that "emerge" from a large system of many "smaller or simpler entities such that ... large entities" [i.e., macroscopic behaviors] arise which "exhibit properties the smaller/simpler entities do not exhibit." [1]. In this thesis we investigate mechanisms and manifestations of emergence in four dynamical systems consisting of many nonlinearly evolving units. These four systems are as follows.

(a) Motion of point particles in a slowly varying Hamiltonian system.

(b) Oscillatory neural networks and jet-lag recovery.

(c) Firing dynamics of integrate-and-fire neuronal networks with inhibitory neurons.

(d) Reservoir computing networks as model-free observers.

We now give brief summaries of our work on these four problems.

1.2 Motion of Point Particles in a Slowly Varying Hamiltonian System

In Chapter 2 we consider the motion of point particles in a particular type of one-degree-of-freedom, slowly changing, temporally periodic Hamiltonian. Through most of the time cycle, the particles conserve their action, but when a separatrix is approached and crossed, the conservation of action breaks down, as shown in previous theoretical studies [2–13]. These crossings have the effect that the numerical solution shows an apparent contradiction. Specifically, we consider two initial constant energy phase space curves $H = E_A$ and $H = E_B$ at time $t = 0$, where H is the Hamiltonian and E_A and E_B are the two initial energies. The curve $H = E_A$ encircles the curve $H = E_B$. We then sprinkle many initial conditions (particles) on these curves and numerically follow their orbits from $t = 0$ forward in time by one cycle period. At the end of the cycle the vast majority of points initially on the curves $H = E_A$ and $H = E_B$ now appear to lie on two new constant energy curves $H = E'_A$ and $H = E'_B$, where the B' curve now encircles the A' curve (as opposed to the initial case where the A curve encircles the B curve). Due to the uniqueness of Hamilton dynamics, curves evolved under the dynamics cannot cross each other. Thus the apparent curves $H = E'_A$ and $H = E'_B$ must only be approximate representations of the true situation that respects the topological exclusion of curve

crossing. In Chapter 2 we resolve this apparent paradox and study its consequences. For this purpose we introduce a 'robust' numerical simulation technique for studying the complex time evolution of a phase space curve in a Hamiltonian system. In Sec. 2.3 and Sec. 2.4 we also consider how a very tiny amount of friction can have a major consequence, as well as what happens when a very large number of cycles is followed. We also discuss in Sec. 2.5 how this phenomenon might extend to chaotic motion in higher dimensional Hamiltonian systems.

1.3 Dynamics of Three Types of Neural Networks

In Chapters 3 to 5 we study three types of neural networks and their dynamics. They are oscillatory neural networks, integrate-and-fire neural networks, and a type of artificial recurrent neural network called a "reservoir computing network" (RCN).

1.3.1 Oscillatory Neural Networks and Jet-lag Recovery

In Chapter 3, using an approximate model of an oscillatory neural network, we study the recovery process of circadian rhythm from jet-lag. Cells in the brain's Suprachiasmatic Nucleus (SCN) are known to regulate circadian rhythms in mammals [14, 15]. We model synchronization of SCN cells using the forced Kuramoto model, which consists of a large population of coupled phase oscillators (modeling individual SCN cells) with heterogeneous intrinsic frequencies and external periodic forcing. Here, the periodic forcing models diurnally varying external inputs such as sunrise, sunset, alarm clocks, etc. We reduce the dimensionality of the system using

the ansatz of Ott and Antonsen [16,17] and then study the effect of a sudden change of clock phase to simulate cross-time-zone travel. We estimate model parameters from previous biological experiments. By examining the phase space dynamics of the model, we study the mechanism leading to the difference typically experienced in the severity of jet-lag resulting from eastward and westward travel [18].

1.3.2 Firing Dynamics of Integrate-and-Fire Neuron Networks with Inhibitory Neurons

In Chapter 4 we study the firing dynamics of a discrete-state/discrete-time version of an integrate-and-fire neuronal network model with both excitatory and inhibitory neurons. The integrate-and-fire neural network is a typical biological neuron model which captures the nature of neuron firing. When the integer-valued state of a neuron exceeds a threshold value, the neuron fires, sends out state-changing signals to its connected neurons and then returns to the resting state. In this model, a continuous phase transition from non-ceaseless firing to ceaseless firing is observed. At criticality, power-law distributions of avalanche size and duration with the previously derived exponents, $-3/2$ and -2, respectively, are observed [19–27]. Using a mean-field approach, we show analytically how the critical point depends on model parameters. Our main result is that the combined presence of both inhibitory neurons and integrate-and-fire dynamics greatly enhances the robustness of critical power law behavior (i.e., there is an increased range of parameters, including both sub- and super-critical values, for which several decades of power law behavior

occurs).

1.3.3 Reservoir Computing Networks as Model-Free Observers

Reservoir computing networks (RCN's) [28] are artificial neural networks with an input layer feeding into a reservoir layer which is a recurrent neural network (RNN) which then connects to an output layer. The advantage of RCN's over traditional RNN's is that training is very computationally cheap. Specifically, training is only done on the output layer, while the connections within the reservoir and the input layer are typically initially determined in a random manner and remain unchanged during training process. In Chapter 5 we explore the potential for RCN's to perform as a universal model-free "observer". Specifically, we consider the case in which a model of the system is unavailable or insufficiently accurate, but "training" time series data of the desired state variables are available for a short period of time, and a limited number of other system variables are continually measured. We propose a solution to this problem using networks of neuron-like units known as "reservoir computers." The measurements that are continually available are input to the network, which is trained with the limited-time data to output estimates of the desired state variables. We demonstrate our method, which we call a "reservoir observer", using the Rössler system and the Lorenz system. Subject to the condition of observability (i.e., whether it is in principle possible, by any means, to infer the desired unmeasured variables from the measured variables), we show that the reservoir observer can be a very effective and versatile tool for robustly reconstructing

unmeasured dynamical system variables.

Chapter 2: Apparent Topologically Forbidden Interchange of Energy Surfaces Under Slow Variation of a Hamiltonian

This chapter is based on work contained in the publication: Apparent topologically forbidden interchange of energy surfaces under slow variation of a Hamiltonian, Lu, Z., Jarzynski, C., and Ott, E., Phys Rev E, 91(5), 052913 (2015),

2.1 Introduction

2.1.1 Background

The principle of adiabatic invariance is a key concept for Hamiltonian systems with one degree of freedom that depend slowly on time. According to this principle, under suitable circumstances, a system's energy will continually change so as to approximately maintain a constant action, $J = \oint p \cdot \mathrm{d}q$, in the current temporally frozen Hamiltonian. This approximation applies when the time scale for change in the Hamiltonian sufficiently exceeds the orbital period of the particle in the current frozen Hamiltonian.

Recently, a particular protocol for the slow, cyclic variation of a Hamilto-

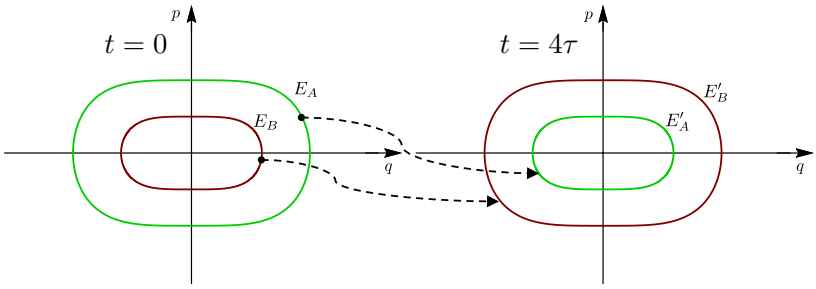


Figure 2.1: Apparent (not real) interchange of energy surfaces.

nian with one degree of freedom was proposed [29], which appears to have counter-intuitive behavior. Reference [29] was originally motivated by earlier work on micro-canonical "Szilard engines" [30] which exhibit the same counter-intuitive behavior and which have been explored further in Ref. [31]. The proposed Hamiltonian was of the form,

$$H(q, p; \vec{\lambda}) = \frac{1}{2}p^2 + V(q, \vec{\lambda}), \tag{2.1}$$

with the time variation of $\vec{\lambda}(t)$ slow, and the parameter vector $\vec{\lambda}$ returning to its initial value after one cycle, as specified in detail in Sec. 2.1.2. At the beginning of the cycle, imagine that a large number N_A of initial conditions are spread uniformly on the energy curve, $H(q, p; \vec{\lambda}(0)) = E_A$, and a second large number N_B are spread uniformly on another energy curve, $H(q, p; \vec{\lambda}(0)) = E_B$, where $E_A > E_B$, and therefore the curve $H = E_A$ encircles the curve $H = E_B$. Numerically, following the Hamiltonian dynamics of all these points through one cycle $\vec{\lambda}(0) = \vec{\lambda}(4\tau)$ (for later notational convenience we denote the cycle duration as 4τ), it is found [29] that at the end of the cycle, to a very good approximation, the vast majority of A points have the same final energy E'_A, which is smaller than the corresponding approximate energy E'_B of the vast majority of B points. Thus $E'_A < E'_B$ and the

curve $H = E'_B$ encircles the curve $H = E'_A$ (see Fig. 2.1). Neglecting a very small fraction of points (which gets smaller as τ increases), it appears that the A and B points approximately lie on constant energy curves that have interchanged; that is, at $t = 0$ the curve of A points encloses the curve of B points, while at $t = 4\tau$ the curve of B points appears to enclose the curve of A points (see Fig. 2.1).

The situation described above shows that the action, which is adiabatically invariant under appropriate conditions, is changed through one cycle of the protocol ($J(E'_A) \neq J(E_A)$, $J(E'_B) \neq J(E_B)$). As discussed in Sec. 2.1.2 and Refs. [2–13], the explanation for this change of action is that the cyclic protocol for $\vec{\lambda}(t)$ drives the system through separatrix-associated orbits with arbitrarily long time scales, thereby violating a necessary condition for action being adiabaticly invariant.

The main significant feature that we wish to emphasize is that the A and B action curves appear to have interchanged. This point will be the focus of our study. In particular, this result should be surprising because the uniqueness of Hamiltonian orbits implies that orbits cannot cross one another in phase space. Thus a change such as that depicted in Fig. 2.1 is topologically forbidden. Hence, while Fig. 2.1 may apply in some approximate sense, the true situation must be different. Our question is: what is the nature of this difference and its dynamical origin?

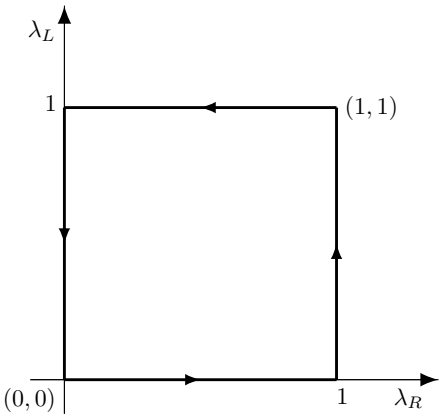


Figure 2.2: The parameter vector $\vec{\lambda} = (\lambda_R, \lambda_L)$ starts from $(0,0)$ when $t = 0$ and reaches $(1,0)$ when $t = \tau$, $(1,1)$ when $t = 2\tau$, $(0,1)$ when $t = 3\tau$, and returns back to $(0,0)$ when $t = 4\tau$. Each of the four phases of the full cycle has duration τ, and $\vec{\lambda}$ changes linearly with time during each of the four phases, Eq. (2.3).

2.1.2 Dynamics of the Hamiltonian

To realize the scenario described in Sec. 2.1.1, we consider a particle with one degree of freedom, described by the parameter-dependent Hamiltonian,

$$H(q,p;\vec{\lambda}) = \frac{p^2}{2} + V(q;\vec{\lambda}) = \frac{p^2}{2} + q^4 - \begin{cases} \lambda_L q^2 & \text{if } q \le 0, \\ \lambda_R q^2 & \text{if } q \ge 0, \end{cases} \tag{2.2}$$

where q is position, p is momentum, and $\vec{\lambda} = (\lambda_R, \lambda_L)$ are parameters that modulate the shape of the potential function $V(q;\vec{\lambda})$. As shown in Fig. 2.2, the time dependent parameter vector $\vec{\lambda}(t)$ proceeds from $(0,0)$ to $(1,0)$ to $(1,1)$ to $(0,1)$ and back to $(0,0)$, while the potential function deforms as shown in Fig. 2.3. More precisely, we

take $\vec{\lambda}$ to vary with time as follows:

$$(\lambda_R, \lambda_L) = \begin{cases} \left(\frac{t}{\tau}, 0\right) & \text{if } 0 < t \leq \tau \quad \text{(phase 1)}, \\ \left(1, \frac{t}{\tau} - 1\right) & \text{if } \tau < t \leq 2\tau \quad \text{(phase 2)}, \\ \left(3 - \frac{t}{\tau}, 1\right) & \text{if } 2\tau < t \leq 3\tau \quad \text{(phase 3)}, \\ \left(0, 4 - \frac{t}{\tau}\right) & \text{if } 3\tau < t \leq 4\tau \quad \text{(phase 4)}. \end{cases} \tag{2.3}$$

Since the Hamiltonian system is time-dependent, the energy of the particle changes with time as

$$\frac{dE}{dt} = \frac{\partial H}{\partial q}\dot{q} + \frac{\partial H}{\partial p}\dot{p} + \frac{\partial H}{\partial \vec{\lambda}}\dot{\vec{\lambda}} = \frac{\partial H}{\partial \vec{\lambda}} \cdot \frac{d\vec{\lambda}}{dt}. \tag{2.4}$$

Thus, during phase 1 of the protocol ($0 < t \leq \tau$), the particle loses energy ($dE/dt = -q^2\tau^{-1}$) when it is traveling in the region $q \geq 0$; but it conserves its energy when $q < 0$. During phase 2 it loses energy when $q < 0$ and conserves energy when $q \geq 0$. During phase 3, the particle gains energy ($dE/dt = q^2\tau^{-1}$) when $q > 0$ and conserves energy when $q \leq 0$. During phase 4, it gains energy when $q < 0$ and conserves energy when $q \geq 0$. Upon completing the protocol, the energy of a particle is typically substantially changed from its initial value.

We choose two initial ($t = 0$) curves in (q, p) phase space: curve A and curve B defined as Hamiltonian level sets $H = E_A$, and $H = E_B$ (Figs. 2.3(a) and 2.4(a)), where, corresponding to these energies, the curves enclose areas in (q, p) phase space

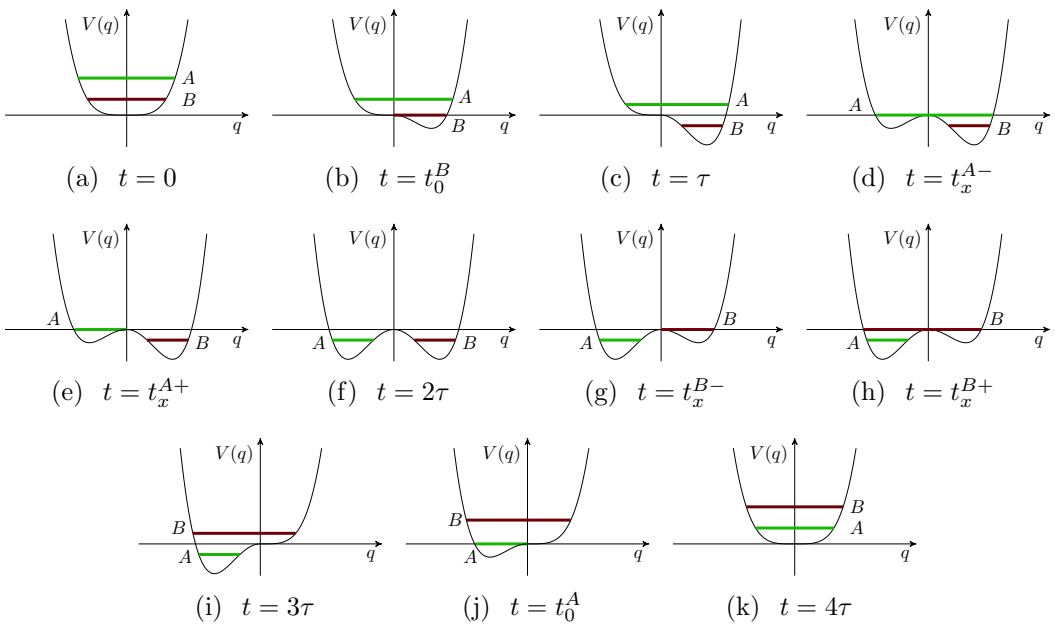


Figure 2.3: Snapshots at $t = 0$, t_0^B, τ, t_x^{A-}, t_x^{A+}, 2τ, t_x^{B-}, t_x^{B+}, 3τ, t_0^A, and 4τ. The green (light gray) and red (dark gray) horizontal lines represent Set A and Set B. Initially, Set A is above Set B, where the height represent the energy of the ensemble. At $t = 4\tau$, the Set B is above Set A.

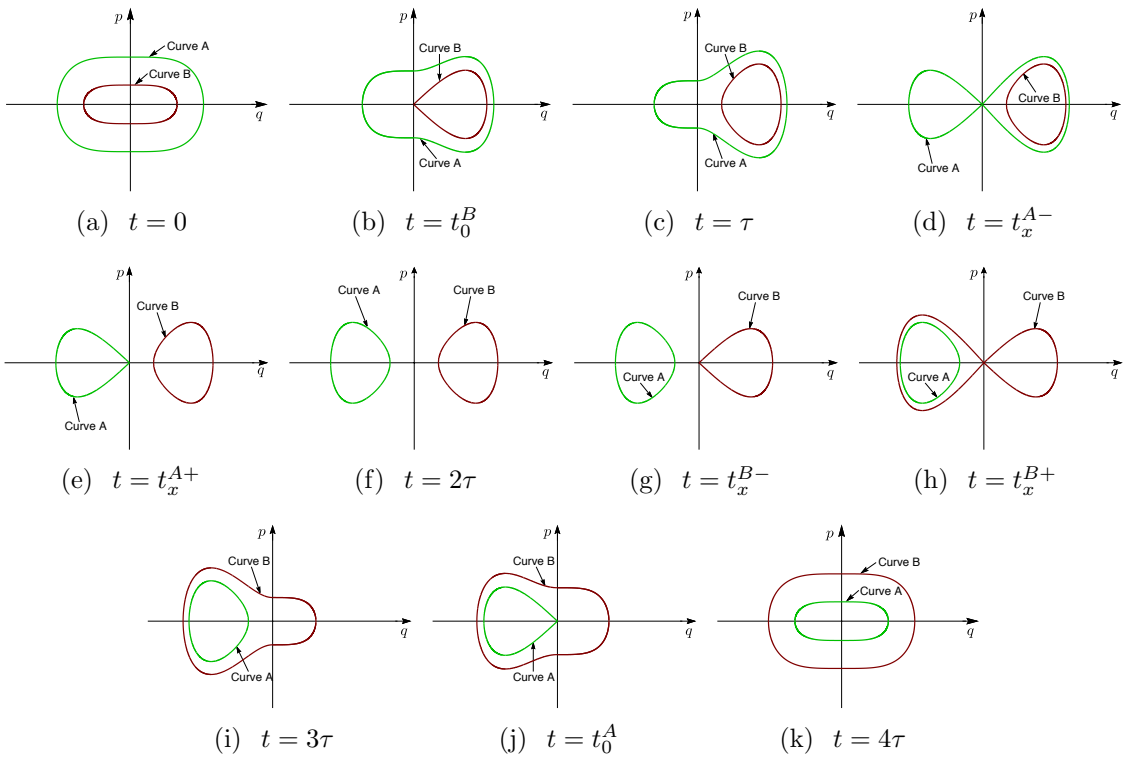
(actions) J_A and J_B that satisfy

$$0 < J_B < J_x < J_A < 2J_x. \tag{2.5}$$

Here J_x is the action enclosed by one lobe of the figure-8 shaped separatrix curve $H(q, p, \vec{\lambda} = (1, 1)) = 0$, or $p^2/2 + q^4 - q^2 = 0$. Thus

$$J_x = 2\sqrt{2} \int_0^1 \sqrt{q^2 - q^4} dq = 2\sqrt{2}/3. \tag{2.6}$$

Considering a cycle of our protocol (Fig. 2.2), we now *define* a hypothetical adiabatic evolution for these phase space curves. This defined evolution is illustrated in Fig. 2.3 and 2.4. At all times curves A and B are curves of constant H. During



(a) $t = 0$ (b) $t = t_0^B$ (c) $t = \tau$ (d) $t = t_x^{A-}$

(e) $t = t_x^{A+}$ (f) $t = 2\tau$ (g) $t = t_x^{B-}$ (h) $t = t_x^{B+}$

(i) $t = 3\tau$ (j) $t = t_0^A$ (k) $t = 4\tau$

Figure 2.4: Assumed evolution of curve A and curve B when $t = 0$, t_0^B, τ, t_x^{A-}, t_x^{A+}, 2τ, t_x^{B-}, t_x^{B+}, 3τ, t_0^A, and 4τ.

phase 1 of our protocol (Figs. 2.3(a) to 2.3(c) and 2.4(a) to 2.4(c)), the energies corresponding to both curves decrease so as to conserve their enclosed actions (the adiabatic invariant). As shown in Figs. 2.3(b) and 2.4(b) there will be some time $t_0^B \in (0, \tau)$ when the energy of curve B changes sign, $E_B(t = t_0^B) = 0$. When $t = \tau$, under the condition given by Eq. (2.5), curve B is trapped in the $q > 0$ well of the potential function, while curve A is not (Figs. 2.3(c) and 2.4(c)).

During phase 2 of our protocol, the energy of curve B is constant, because the potential V remains fixed in the region $q \geq 0$. For $\tau < t < t_x^A < 2\tau$, the energy of curve A decreases to conserve its enclosed action. The time t_x^A (which corresponds to the pseudo-crossing time defined in Ref. [6]) is defined as the time at

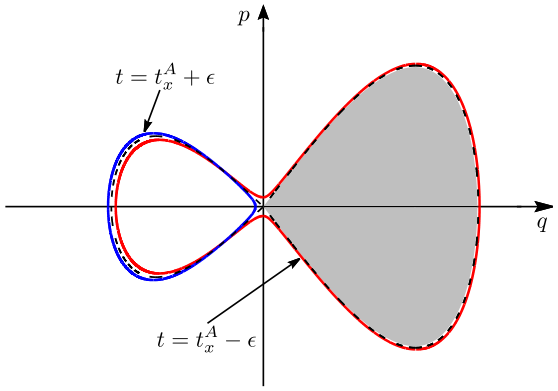


Figure 2.5: Discontinuity of the adiabatic action invariant of A particles at time t_x^A in the hypothetical evolution.

which constancy of the action enclosed by curve A implies that the energy of curve A becomes $V(q = 0) = 0$, which corresponds to the dashed separatrix curve in Fig. 2.5. Motivated by the fact that particles lose energy when $q < 0$ (but not when $q > 0$) during phase 2, we define our hypothetical evolution to be such that curve A is located in $q < 0$ for $t_x^A < t < 2\tau$. This is illustrated in Figs. 2.4(e) and 2.4(f), and by the blue (dark gray) curve in Fig. 2.5. By this definition the energy of particles on curve A evolves continuously in time, but there is a discontinuity in the action enclosed by curve A, by the amount $J_x = 2\sqrt{2}/3$ as t crosses t_x^A [6]. Specifically, the action associated with curve A decreases discontinuously by the area of the right lobe of the $H(q, p; \vec{\lambda}(t_x^A)) = 0$ separatrix depicted as the gray shaded area in Fig. 2.5 (see also Figs. 2.3(d,e) and 2.4(d,e) where t_x^{A-} and t_x^{A+} denote time instants just before and just after t_x^A). For $t_x^A < t \leq 2\tau$, the enclosed action area of curve A is conserved. At the end of phase 2 ($t = 2\tau$), the two curves are trapped in different wells (Figs. 2.3(f) and 2.4(f)). Curve B is trapped in the right well ($q > 0$) with action J_B, and curve A is trapped in the left well with action $J_A - J_x$.

During phase 3 of the protocol, for $2\tau < t < t_x^B$, the potential function increases on the $q > 0$ side and the energy of curve B increases maintaining constant action, becoming larger than the energy of curve A. Again, there is a time $t_x^B \in [2\tau, 3\tau]$ at which the energy of curve B reaches the value $V(q = 0) = 0$, and there is then a discontinuous increase in the action enclosed by curve B by the amount $J_x = 2\sqrt{2}/3$, similar to what is illustrated in Fig. 2.5 but with $q \to -q$ and the time ordering reversed; see Figs. 2.3(g,h) and 2.4(g,h). For all subsequent time $t_x^B < t < 4\tau$ the enclosed actions of curves A and B remain constant (Figs. 2.3(i,j,k), 2.4(i,j,k)), with values $J_A - J_x$ and $J_B + J_x$.

During phase 4, the left potential well rises and the potential returns to its original quartic shape. We use $t_0^A \in (3\tau, 4\tau)$ to denote the time when the energy of curve A changes sign from negative to positive; see Figs. 2.3(j) and 2.4(j). Note that at $t = 4\tau$, the energy of curve B is higher than that of curve A (Figs. 2.3(k) and 2.4(k)). Thus, proceeding through one cycle of the protocol, we have flipped the relative values of the energies and actions of the two curves. This flip is due to the two separatrix-crossing-induced action discontinuities: one at t_x^A decreasing the action of curve A, and the other at t_x^B increasing action for curve B. Specifically,

$$J_A(4\tau) = J_A(0) - J_x, \tag{2.7a}$$

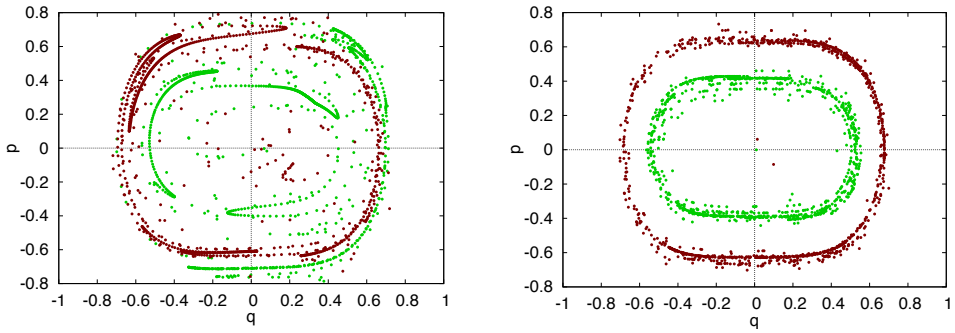
$$J_B(4\tau) = J_B(0) + J_x, \tag{2.7b}$$

where $J_x = 2\sqrt{2}/3$ (see Eq. (2.6)) is the (q, p) area inside the right (left) separatrix lobe at $t = t_x^A$ ($t = t_x^B$). As shown by Eqs. (2.7), the initial A and B curves

exactly interchange under the hypothetical evolution when $J_A(0) - J_B(0) = J_x$. While we could consider other values of $J_A(0) - J_B(0)$, for specificity, we have taken $J_A(0) - J_B(0) = J_x$ in Figs. 2.3-2.5.

As already noted (Sec. 2.1.1) and discussed in more detail in Sec. 2.2, the hypothetical evolution illustrated in Figs. 2.3-2.5 cannot be entirely correct, as it violates the uniqueness and existence theorem of Hamiltonian dynamics. However, the analysis of Cary, Escande and Tennyson [6] implies that this hypothetical evolution is, in a suitable sense, an approximation to the real dynamics. Specifically, in the limit of a very large number of initial conditions sprinkled on curves A and B, and a suitably large value of τ, if we remove from consideration a small fraction f of the particles, then the energies of the remaining particles never deviate from the energies resulting from our defined ideal evolution by more than some small quantity δ, and we can decrease f and δ with increasing cycle time so that $f \to 0$, $\delta \to 0$ as $\tau \to \infty$. Indeed, the results of Ref. [6] imply that f decreases exponentially with τ.

These conclusions are supported by numerical evidence. We simulated 1000 Hamiltonian trajectories with initial conditions sampled from the energy surface $E_A = 0.2401$, and another 1000 trajectories with initial conditions sampled from $E_B = 0.0531$. Fig. 2.6 shows the final conditions of these trajectories at the end of one cycle $t = 4\tau$ for different values of τ. At low τ (Fig. 2.6(a) for $\tau = 20$) the hypothetical curve evolution is seen to very poorly predict the result of the true dynamics. However, as τ is increased (Figs. 2.6(b-d) for $\tau = 100, 500, 1000$) the hypothetical evolution provides an increasingly accurate description of the dynamics. Specifically, for Figs. 2.6(b-d), there is a dense set of A particles colored in green



(a) $4\tau = 80$ (b) $4\tau = 400$

(c) $4\tau = 2000$ (d) $4\tau = 4000$

Figure 2.6: Final result at $t = 4\tau$ from regular simulation with different slowness parameters τ.

(light gray) that seem to be tracing out a constant energy curve of energy E'_A that agrees well with the prediction of Eq. (2.7a), as well as a dense set of B particles colored in red (dark gray) that seem to be tracing out a constant energy curve of energy $E'_B > E'_A$, where the value of E'_B is predicted by Eq. (2.7b). Furthermore, in agreement with Ref. [6], we see that the E'_A and E'_B "curves" in Figs. 2.6(b-d) have a width that decreases with increased τ implying that $\delta \to 0$ as $\tau \to \infty$. Using more sample points, we always see a small number of green (light gray) and red (dark gray) particles that are far from the constant energy E'_A and E'_B curves; we call such particles "delinquents". For instance, in Figs. 2.6(a-b) some red (dark gray) points appear well within the interior of the region encircled by the $H = E'_B$ curve. In

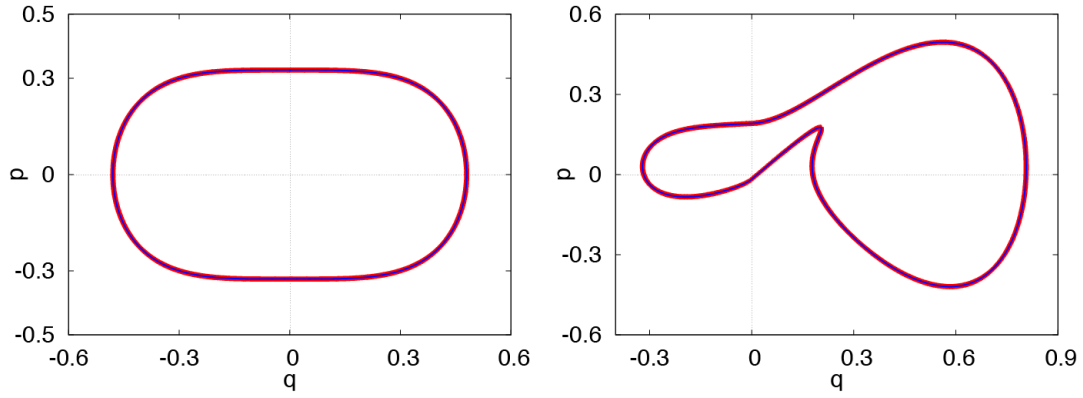
accord with Ref. [6], we see that the fraction of delinquents in Figs. 2.6 gets smaller with increasing τ (i.e., $f \to 0$ as $\tau \to \infty$).

2.1.3 Outline

The rest of this chatper is organized as follows. Section II discusses the resolution of the topological issue raised above, that is, the interchange of the A and B curves, and evidence for the type of asymptotic validity ($f \to 0$, $\delta \to 0$ as $\tau \to \infty$) of the curve dynamics illustrated in Figs. 2.6. This is done partly by our introduction of a 'robust' numerical simulation technique for studying the complex time evolution of a phase space curve in a Hamiltonian system. Section III notes that the adiabatic behavior for $\tau \gg 1$ is dramatically changed with a small amount of friction, and discusses the nature of this change. Section IV shows numerically the results of repeated cycling of this Hamiltonian system, which we relate to a theorem shown in Ref. [13]. Section V discusses how our work might be extended to chaotic motion in higher dimensional systems and summarizes our conclusions.

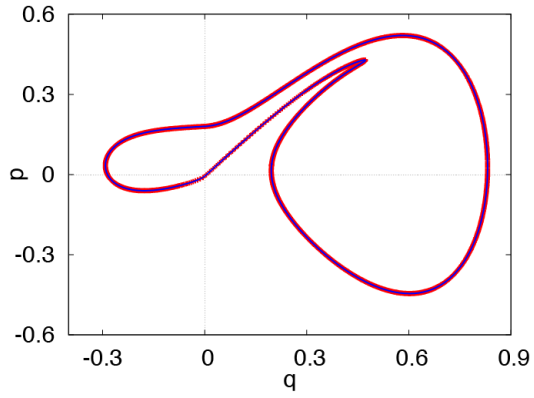
2.2 Resolving the Topology Issue

As discussed in Sec. 2.1 and illustrated in Fig. 2.6, for sufficiently large τ numerical simulations appear to be in agreement with the hypothetical evolution shown in Figs. 2.3 and 2.4, in which the curves A and B evolve through a sequence of action sets of the Hamiltonian. However, this defined dynamics cannot exactly correspond to the real dynamics in this slowly varying Hamiltonian since it violates

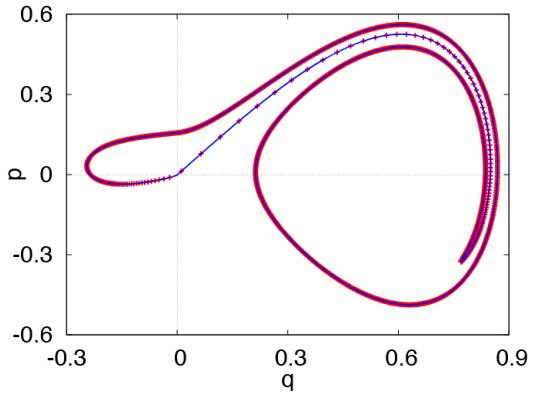


(a) $t = 0$

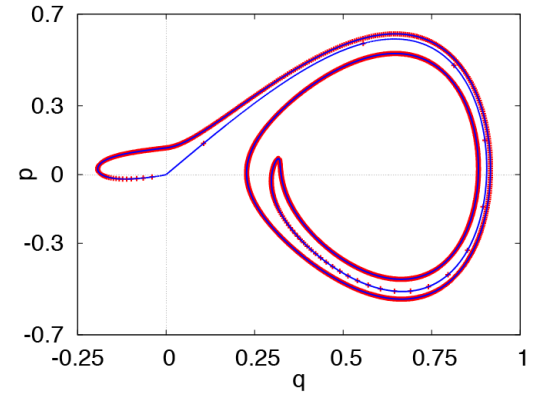
(b) $t = 0.65\tau$



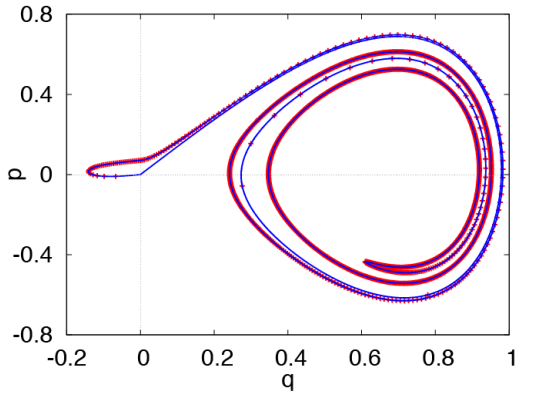
(c) $t = 0.7\tau$



(d) $t = 0.77\tau$

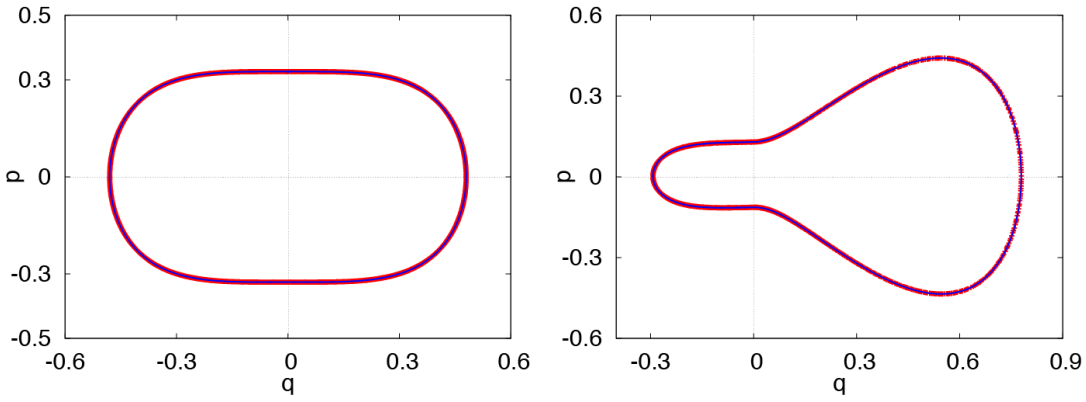


(e) $t = 0.86\tau$



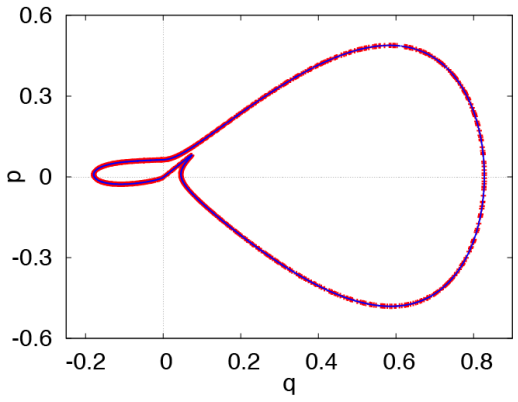
(f) $t = \tau$

Figure 2.7: Six figures are snapshots in phase space showing the evolution of sample points along a positive energy curve at $t = 0$ to the resulting curve at $t = \tau = 20$, from both the regular and the robust simulations. The regular simulation evoloves only those 2000 initial phase points. Its result is shown by the thick red (light gray) "curve" which is made up of 2000 red (light gray) points. The thin blue (dark gray) curve is the result from the robust simulation starting initially with 2000 points.

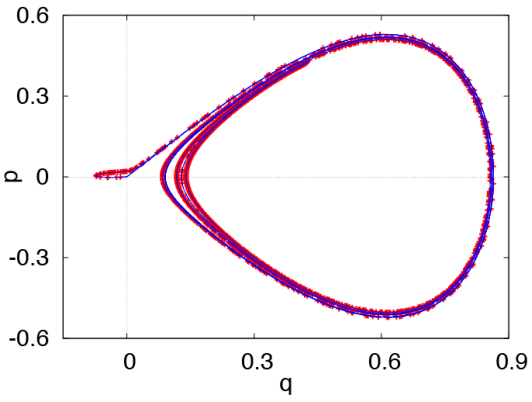


(a) $t = 0$

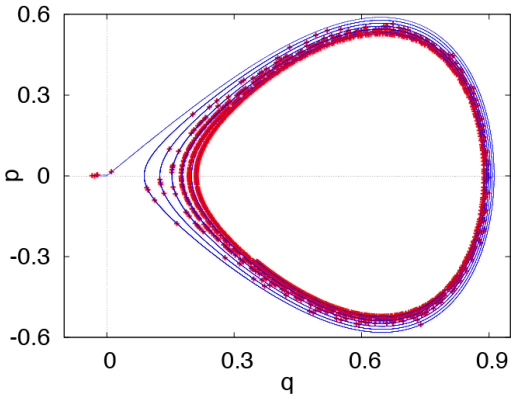
(b) $t = 0.6\tau$



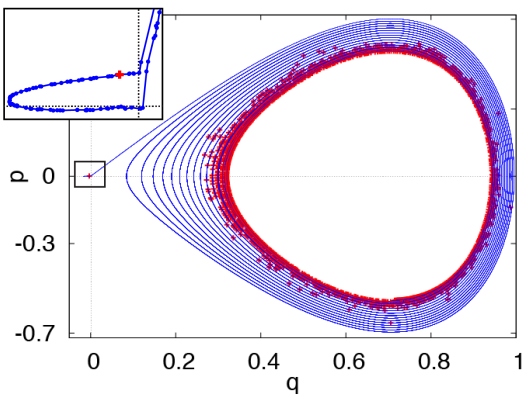
(c) $t = 0.68\tau$



(d) $t = 0.75\tau$



(e) $t = 0.84\tau$



(f) $t = \tau$

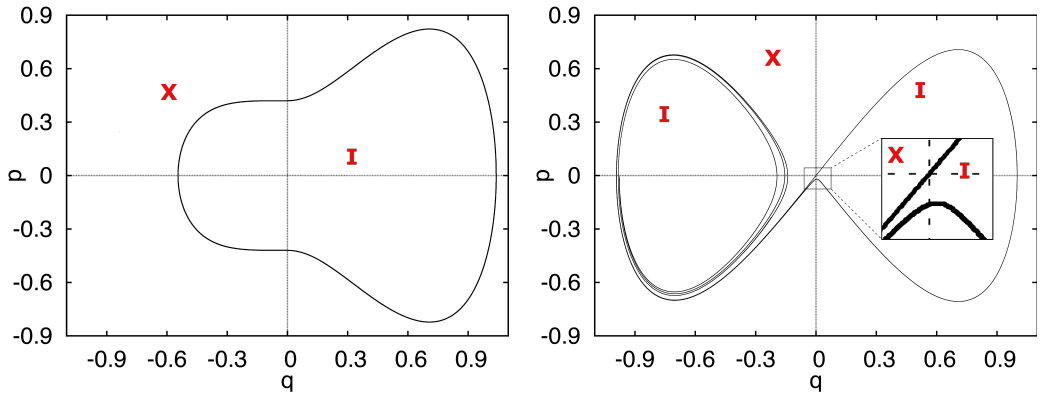
Figure 2.8: Snapshots for $0 < t < \tau$ with a larger $\tau = 200$. Red (light gray) points are results from regular simulation and the blue (dark gray) thin curve is of robust simulation with $\epsilon = 0.0531$. The inset in figure (f) shows a magnification of the small rectangular near the origin, verifying that the origin $(q, p) = (0, 0)$ is still enclosed by the true time-evolved curve at time $t = \tau$.

two topological constraints. E.g., see Ref. [6] for related discussion.

Most obviously, curve A encloses curve B at the beginning, yet curve A is enclosed by curve B at the end of the protocol. If this reflected the true dynamics, then every phase space orbit on the A curve would at some time have to cross an orbit on the B curve, which contradicts the uniqueness of solutions of Hamiltonians. (If there were an orbit crossing time, then at this time an A particle and a B particle would have the same (q, p) value; then by evolving Hamiltonian's equations forward in time or backward in time from this crossing point, we would infer that the two orbits were identical for all time, thus contradicting the supposed existence of a crossing.) This violation of topology is visually evident in Fig. 2.4, as one proceeds from Fig. 2.4(c) to 2.4(i).

A less obvious topological violation occurs between $t = 0$ (Fig. 2.4(a)) and $t = \tau$ (Fig. 2.4(c)) and again between $t = 3\tau$ (Fig. 2.4(i)) and $t = 4\tau$ (Fig. 2.4(k)). This violation arises because the origin $(q, p) = (0, 0)$ is a fixed point; since $\partial V(q, t)/\partial q = 0$ at $q = 0$ for all time, a particle positioned exactly at $(p, q) = (0, 0)$ at $t = 0$ would remain there throughout the cycle. Thus, under the true Hamiltonian dynamics, neither curve A nor curve B could pass through $(q, p) = (0, 0)$, as they do several times during the hypothetical dynamics indicated in Figs. 2.4 and 2.5. The basic discrepancy between the hypothetical and the real evolution can be traced to the instants when curve A or B seemingly "crosses" the origin, as the condition of adiabaticity is then violated due to the fact that the orbit period in the frozen Hamiltonian approaches infinity.

In order to study the apparent "crossing" of $(q, p) \equiv (0, 0)$, we consider the



(a) $t = \tau$ (b) $t = 2\tau$

Figure 2.9: Energy curve dynamics from a robust simulation illustrate the real dynamics of the separatrix "crossing" in the 2nd phase. The insert shows a magnification about the origin for (b). The "X" and "I" symbols label regions that are exterior and interior to the closed curve; these designations are also used for Figs. 2.10-2.11.

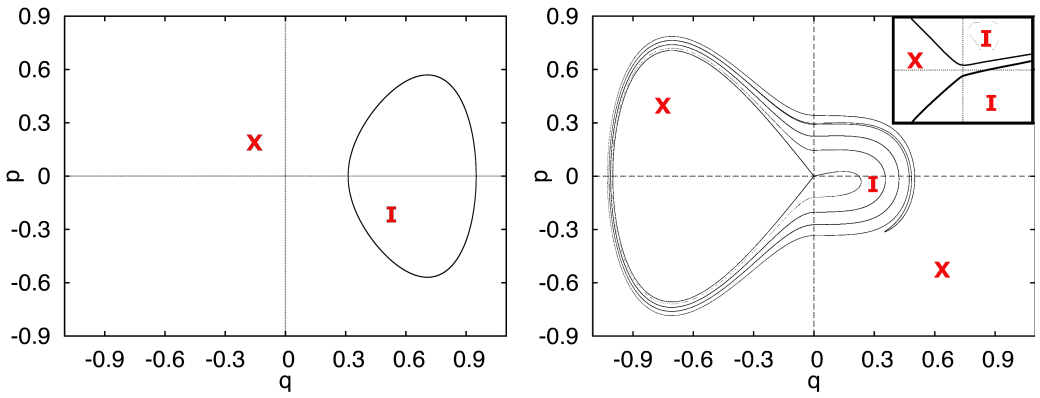
true dynamics when t approaches t_0^A or t_0^B. The part of the curve that is nearer to the origin moves slower than the part that is farther from the origin, since the origin is a fixed point of the dynamics. As a result, instead of crossing the origin, this curve is stretched and wound around the separatrix as it approaches the origin (see subsequent discussion of the blue (dark gray) curves in Figs. 2.7, 2.8, and inset to Fig. 2.8(f)). The situation is somewhat similar when t approaches t_x^A or t_x^B, the difference being that the curve will be stretched and wound around, not the origin, but the separatrix lobe on the side where the potential function is stationary. In the simulations shown in Figs. 2.6, we uniformly put a finite number of sample particles on the initial curves and evolve them in the Hamiltonian system. The sample particles on the stretched portions of the curve eventually become sparse, and, if τ is too long, for any fixed numbers of points placed on the initial curves, some of them become too far separated to form a good approximation to the evolved

"true" curve. Thus the true curves are eventually not resolved by these points, which are shown as red (light gray) dots in Figs. 2.7, 2.8. This is why it appears that the evolved curves cross the origin and each other in the simulation. We will refer to a simulation employing a fixed number of sample points as a regular simulation.

In order to better reveal the true curve dynamics and show how a curve is wound around the origin or a separatrix lobe, we use a simulation method that we call "robust". In contrast to the regular simulation, the robust simulation monitors the distance between each pair of points that occur consecutively along the initial curve. When their separation d becomes larger than a critical value ε we add a new phase point halfway between them. The thin blue (dark gray) curves in Figs. 2.7, 2.8 were obtained in this manner. Note that when part of the curve approaches the origin, it will experience greater curve stretching, and more phase points will be added to that part of the curve. Thus, in the robust simulations the curve made by joining consecutive phase points will always give a good approximation to the true curve. The smaller the value of ε, the less sparse the curve. However, with too small ε, the simulation will require a large population of phase points. Thus, ε should be small enough that the curve is well-resolved but allow reasonable computational cost. We also note that the number of phase points that are required by the robust simulations increases with increasing τ due to increase in the number of curve folds (compare Figs. 2.7 for $\tau = 20$ with Figs. 2.8 for $\tau = 200$).

During the first phase of the protocol, curve B approaches the origin when $t = t_0^B$. Figs. 2.7 and 2.8 are snapshots of regular and robust evolution of energy curve B for $t \in [0, \tau]$ with $\tau = 20$ and $\tau = 200$, respectively. The red (light gray)

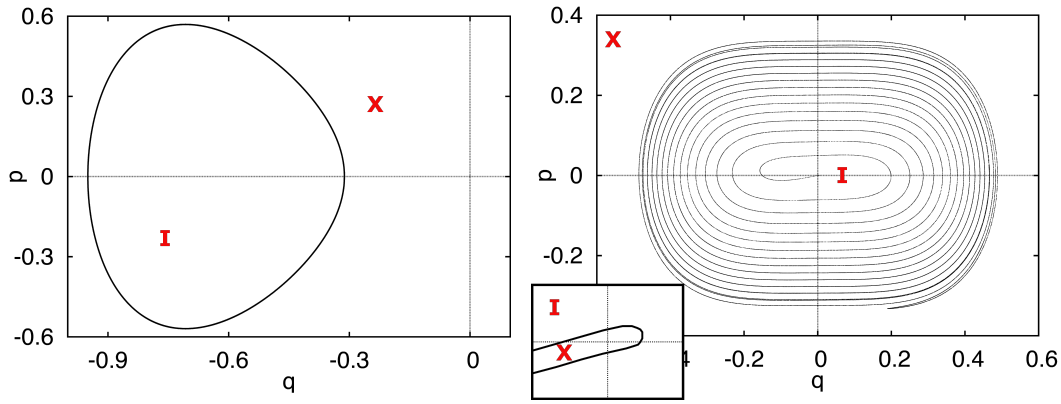
dots from the regular simulation do not maintain a continuous curve, while the blue (dark gray) plots from the robust simulation do. It is clear that the true curve does not cross the origin, but is wound around it, as shown by the blue (dark gray) curve in Figs. 2.7 with $\tau = 20$ and the inset to Fig. 2.8(f). The same multi-foil structure as in Figs. 2.7, but with more layers, appears in Figs. 2.8 where $\tau = 200$. We conclude that (i) the area enclosed by the true deformed curve is conserved (by Liouville's theorem); (ii) the multi-foil structure has more layers with larger τ; (iii) delinquents of the regular simulation are a small portion of phase points lying on the stretched foils (Fig. 2.8(f)). For larger τ, the curve experiences greater stretching, while the proportion of delinquents becomes smaller, and most of the phase points stay near the curve predicted by the hypothetical evolution.



(a) $t = 2\tau$ (b) $t = 3\tau$

Figure 2.10: Energy curve dynamics from robust simulation illustrates the real dynamics of the separatrix "crossing" in the 3rd phase. A schematic illustration of a magnification around the origin is shown as the inset in panel (b).

During the second phase ($t \in [\tau, 2\tau]$) in the hypothetical evolution, curve A "crosses" the separatrix orbit lobe on the $q > 0$ side together with the fixed point $(q, p) = (0, 0)$ when $t = t_x^A$. This appears to happen in the regular simulation for



(a) $t = 3\tau$ (b) $t = 4\tau$

Figure 2.11: Energy curve dynamics from robust simulation illustrates the real dynamics of the separatrix "crossing" in the 4th phase, starting with a Hamiltonian level curve of action $J_A(0) - J_x$ at time $t = 3\tau$. The inset shows a schematic illustration of a magnification around the origin of the robust curve calculation result at time $t = 4\tau$.

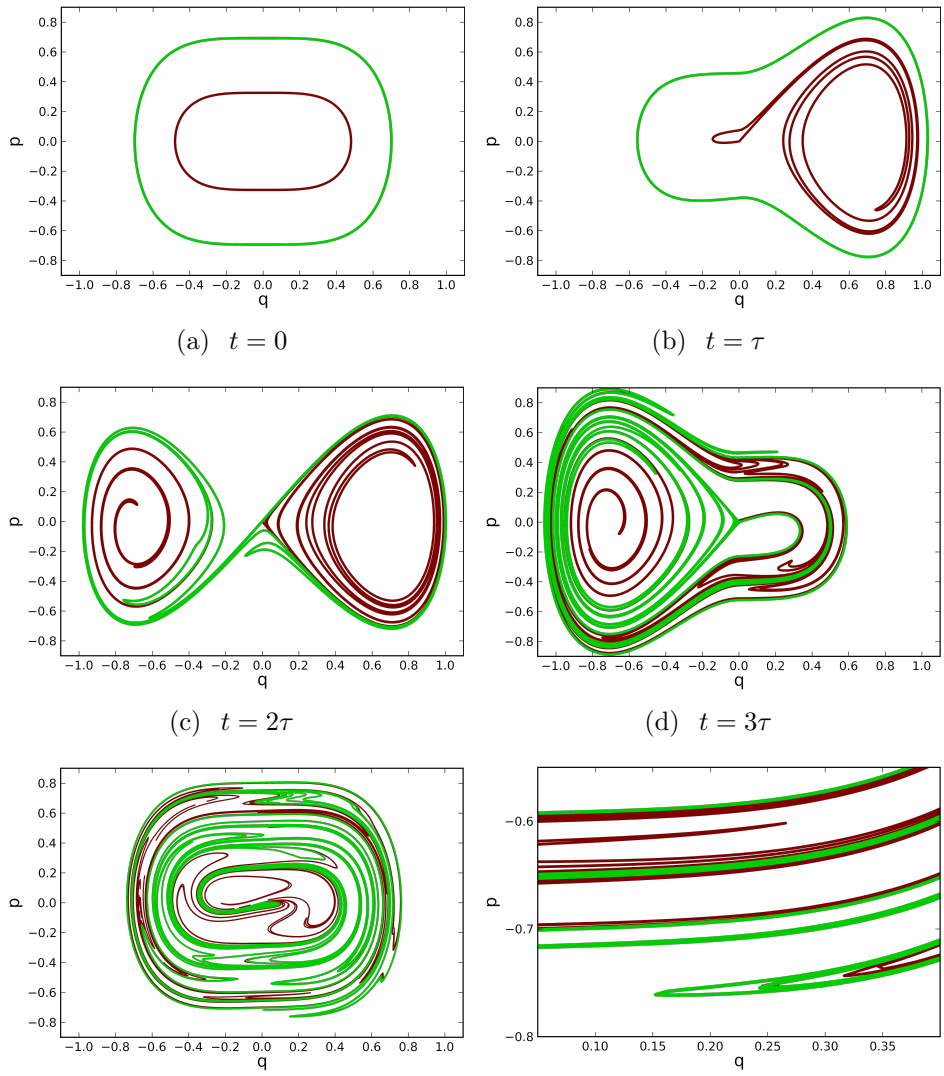
the same reason we discussed for the first phase. A curve made of numerical sample particles from the regular simulation will appear to be broken when consecutive points move far apart as the true curve is stretched and wound around the separatrix lobe. To illustrate the real dynamics, we use the robust simulation for $t \in [\tau, 2\tau]$ and prepare an initial ($t = \tau$) curve A as specified in the hypothetical evolution (Fig. 2.4(c)). In Fig. 2.9(a), the closed curve encloses area J_A that satisfies $J_x < J_A < 2J_x$. At the end of the second phase, the curve, as shown in Fig. 2.9(b), deforms into a curve enclosing both the right lobe of the separatrix orbit with an area J_x and a region on the left side. For $t_x^A < t \leq 2\tau$, the area on the left side that is enclosed by the true curve is well-approximated by the area enclosed by the hypothetical curve. The fraction of particles in $q > 0$ from a regular simulation becomes very small for large τ, and the $q > 0$ lobe of area J_x eventually appears not to be encircled in a regular simulation. However, by Liouville's theorem the area

enclosed by the deformed true curve is conserved. We obtain $J_A = J'_A + J_x$ where J'_A is the A action for the hypothetical evolution for $t > t_x^A$.

During the third phase in the hypothetical evolution (Figs. 2.4(f-i)), curve B "swallows" the left separatrix orbit lobe and the separatrix point (origin) at time $t = t_x^B$ ((Figs. 2.4(g-h))). During this phase of the evolution curve B gains energy, passing from negative to positive energies at $t = t_x^B$. To investigate the real dynamics, we initialized trajectories on the level surface $H = J_B(0)$ at time $t = 2\tau$, which corresponds to the red (dark gray) curve in Fig. 2.4(f). We then evolved this curve under the robust dynamics from $t = 2\tau$ to $t = 3\tau$. The results are shown in Fig. 2.10. The envelope of the curve in Fig. 2.10(b) has the same shape as the hypothetical red (dark gray) curve in Fig. 2.4(i). Most of the sample points from the regular simulation are near the envelope of the curve in Fig. 2.10(b), and there are a small number of delinquents, which we expect will disappear if τ is large enough. Although it is not obvious in the figure, the curve does not enclose the area confined by the left lobe of the separatrix orbit. Instead, it encloses the area confined within the multi-foiled structure that has numerous layers.

During the fourth phase, curve A gains energy and seems to cross the origin in the regular simulation. This situation is similar that of curve B's "crossing" in the first phase, with the direction of the crossing being opposite. Curve A starts from inside the potential well on the right side and moves out at the end ($t = 4\tau$). Figure 2.11 shows how the curve in Fig. 2.11(a) at $t = 3\tau$ evolves by the end of the forth phase, $t = 4\tau$. Notice that the curve becomes stretched and thus very thin. The curve that appears to terminate at the origin in Fig. 2.11(b) is not a single line

but a folded curve enclosing the origin (see inset to Fig. 2.11(b)). In the regular simulation, most of the particles lie close to the outer envelop of the structure shown in Fig. 2.11(b) while a small number of particles, the delinquents, can be found along the highly stretched foils.



(a) $t = 0$ (b) $t = \tau$

(c) $t = 2\tau$ (d) $t = 3\tau$

(e) $t = 4\tau$ (f) $t = 4\tau$ zoom in

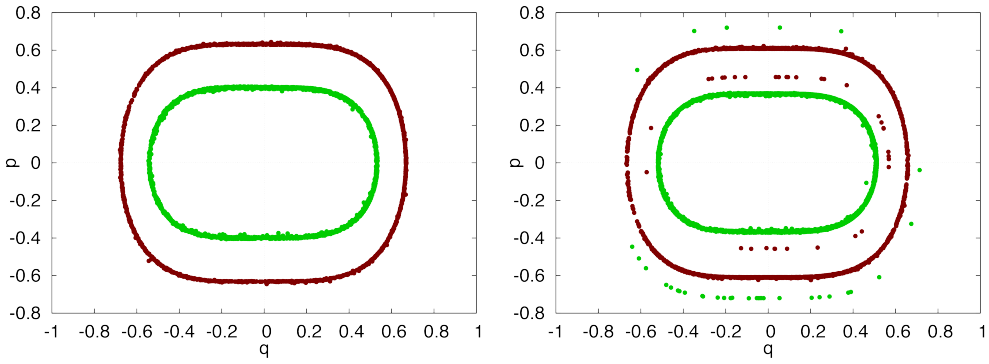
Figure 2.12: Snapshots for the full protocol with $\tau = 20$. Panel (f) is a zoom magnification of a small region of (e).

Over the entire protocol, the curves A and B experience the four deformations described above. Each deformation involves stretching a curve, A or B, as its energy

approaches zero from above or below. The multifoil structures produced by these stretching deformations satisfy the topological constraints imposed by Hamilton's equations, while simultaneously explaining the apparent success of the hypothetical dynamics defined in Sec. 2.1 (Figs. 2.3, 2.4). We have illustrated each deformation separately (Figs. 2.8-2.11), but when the system is subjected to the full protocol the four deformations occur in sequence, giving rise to a convoluted mixing of the curves A and B. When using robust simulations to study the evolution of both curves, we have found that for large τ (e.g., $\tau = 200$) the proliferation of phase points soon becomes excessive and it is numerically infeasible to follow the full evolution through an entire cycle $[0, 4\tau]$. However, for small enough τ we are able to simulate the cycle robustly. We show the results for $\tau = 20$ in Figs. 2.12. Although it is not evident to the eye, the evolving green (light gray) curve A encloses the red (dark gray) curve B at all times, and each curve remains topologically equivalent to a simple, closed loop enclosing the origin $(0, 0)$. A regular simulation is incapable of resolving the stretched, thin structure created by these dynamics.

2.3 Model with Damping

Although the regular dynamics approximately follows the hypothetical evolution when τ is large, adding a small amount of friction can change the evolution



(a) $\gamma = 0$ (b) $\gamma = 2.0 \times 10^{-5}$

Figure 2.13: Result of two energy curves after going through the whole protocol with (b) and without (a) friction, $4\tau = 4000$ and $N = 2000$.

dramatically. In this case we consider the evolution,

$$\dot{q}(t) = p, \tag{2.8a}$$

$$\dot{p}(t) = -\frac{\partial H(q,p,t)}{\partial q} - \gamma p, \tag{2.8b}$$

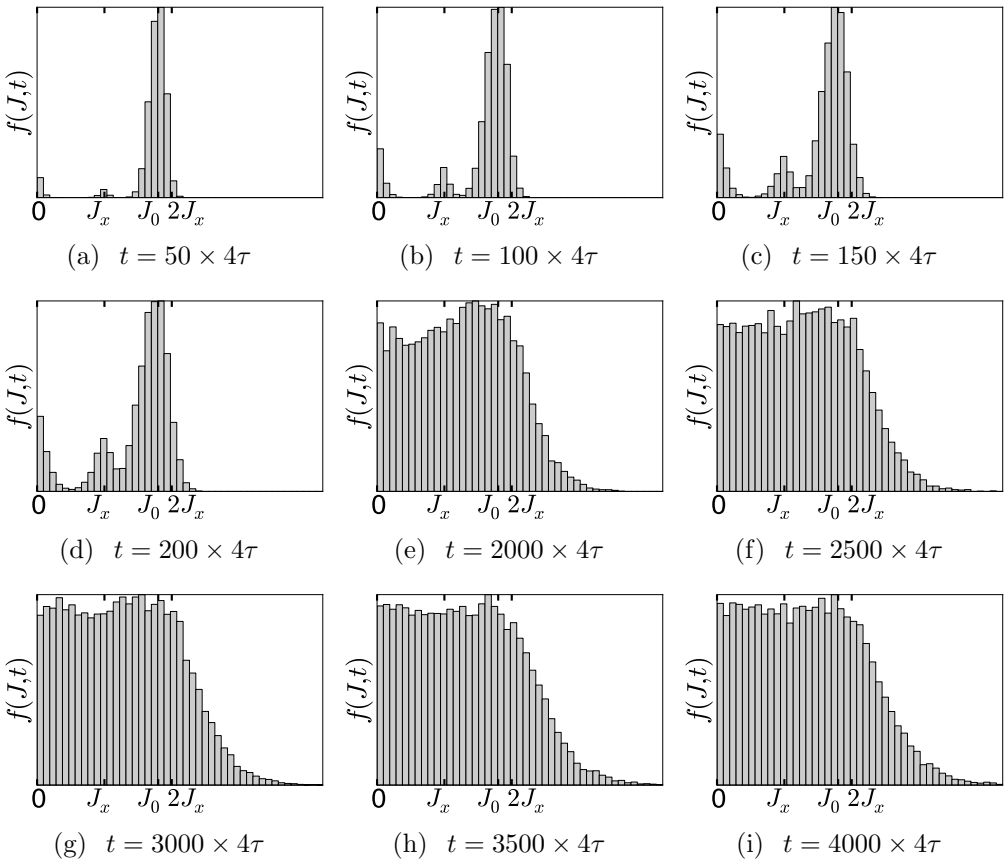
where γ is the friction coefficient. Figures 2.13 show regular simulation results for $\tau = 1000$. Figure 2.13(a) has $\gamma = 0$, while Fig. 2.13(b) has $\gamma = 2 \times 10^{-5}$ (since $4\gamma\tau = 0.08$ the effect of friction over the cycle time is fairly small). As seen in Fig. 2.13(b), by adding friction, $\gamma = 2 \times 10^{-5}$, the result of a regular simulation is changed in an interesting way. Particles now appears to lie on four different energy curves instead of two. The two densest curves are close to the $\gamma = 0$ curves in Fig. 2.13(a), but there are additionally two sets of delinquents that appear to populate two other curves, with approximate actions J_x and $2J_x$. The impact of the friction term is negligible at all times except when t is close to t_x^A and t_x^B, when the seeming separatrix "crossings" occur. To understand this we will discuss the

formation of the curve of green (light gray) delinquents that encircle the dense red (dark gray) curve in Fig. 2.13(b). (Similar considerations explains the presence of the curve of red (dark gray) delinquents occurring between the dense red (dark gray) and green (light gray) curves of Fig. 2.13(b).) To start, we notice that in Fig. 2.5 (for $t \simeq t_x^A$), when the A curve is about to cross the separatrix, it is very close to the right lobe of the separatrix, and its energy is slightly positive. As the trajectories corresponding to this curve gradually lose energy to friction, some of the trajectories in the region $q > 0$ acquire slightly negative energies, causing them to be trapped in the right well. Let us consider the fate of these delinquents during the remainder of the cycle. For $t_x^A < t < 2\tau$, these points remain in the right well – where the potential energy is not changing – and they evolve along approximately closed orbits enclosing an area J_x. At $t = 2\tau$ the potential in $q > 0$ begins to increase, and these trajectories now go over the potential hilltop at $q = 0$ and follow orbits that approximately encircle both the $q < 0$ and the $q > 0$ separatrix lobes. Thus for t slightly greater than 2τ these orbits now encircle an area corresponding to a total action of approximately $2J_x$. As t continues to increase, these orbits continually gain energy while at the same time conserving their action of approximately $2J_x$. At the end of the cycle, these trajectories are located along the curve of green (light gray) delinquents shown in Fig. 2.13(b). Similar considerations explain the "curve" of red (dark gray) delinquents. Thus adding friction not only increases the number of delinquents, but also changes the distribution of sample particles in phase space.

2.4 Repeated Cycling of the Protocol Given by Eqs. (2.2) and (2.3)

2.4.1 Evolution of the Action Distribution Function

As we have seen, after a single cycle of the protocol given by Eqs. (2.2) and (2.3), the predictions of the hypothetical dynamics for curves A and B acquire some error. Although this error is small when τ is large, it can be expected to accumulate if the protocol is repeated over many cycles. The effect of repeated separatrix crossing on the distribution of action has also been studied in Refs. [12,13,32,33]. One of the important results is that in the limit of slow cycling, a Hamiltonian with a pulsating separatrix is chaotic in the region of phase space swept by the separatrix, and that the repeated separatrix crossings lead to diffusion of the adiabatic invariant [13]. To numerically investigate these anticipated effects in our system, we performed slow simulations ($4\tau = 2000$) in which trajectories were followed over 4000 cycles of the protocol. We used $20,000$ trajectories, with initial conditions sampled from a curve enclosing $J_0 = 1.696$, and we monitored the evolution of the probability distribution of the action with time. Fig. 2.14 shows this probability distribution of action after various numbers of cycles, from 50 to 4000. As shown in Fig. 2.14(a), after 50 cycles of repeated protocols, the width of the peak of the simulated action distribution originally located at the initial action value $J = J_0$ expands. From Figs. 2.14(a-d), we also notice two other smaller peaks forming at $J = 0$ and $J = J_x$. This corresponds to the fact that each time a fixed point crossing (separatrix crossing) occurs, the "closed curve" formed by particles winds around the fixed point



(a) $t = 50 \times 4\tau$ (b) $t = 100 \times 4\tau$ (c) $t = 150 \times 4\tau$

(d) $t = 200 \times 4\tau$ (e) $t = 2000 \times 4\tau$ (f) $t = 2500 \times 4\tau$

(g) $t = 3000 \times 4\tau$ (h) $t = 3500 \times 4\tau$ (i) $t = 4000 \times 4\tau$

Figure 2.14: Simulated histograms of action J after evolving the system different numbers of cycles.

(separatrix lobe). Most of particles approximately follow the hypothetical dynamics so that they form a closed curve which seemingly crosses the fixed point (separatrix lobe), but a portion of particles remain close to the fixed point where $J = 0$ (the separatrix lobe where $J = J_x$) after this crossing. This is also seen in Fig. 2.8(f). After the fixed point crossing in the first phase, a red (light gray) particle still remains close to the fixed point. More generally [6], we expect that the number of particles that end up close to the fixed point or the separatrix lobe becomes smaller with larger τ. Since adiabaticity is spoiled by separatrix crossing and by the limited slowness of the simulation, the argument that a particle with initial action $J \in [0, 2J_x]$ should continue to have its action in $[0, 2J_x]$ is only true for most of particles, but not all. The simulated results show that with more and more cycles, an increasing number of particles end up with action exceeding $2J_x$. We see from Figs. 2.14(g), 2.14(h) and 2.14(i) that after 3000 cycles, the distribution becomes flat in $[0, 2J_x]$ by the end of the simulation, while the tail in $J > 2J_x$ keeps growing. The

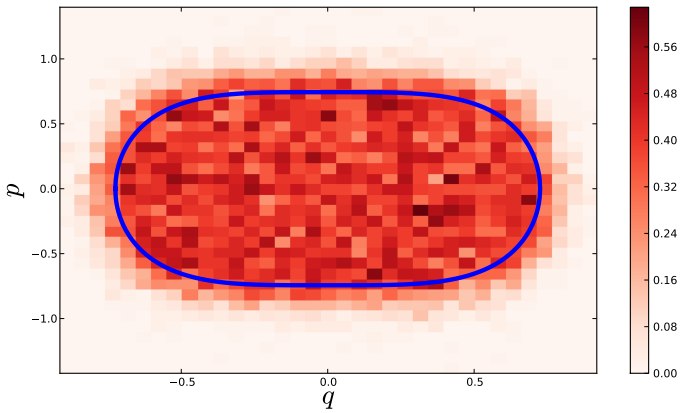


Figure 2.15: Simulated histogram in phase space after 4000 repeating cycles. The curve $J = 2J_x$ is shown in blue (dark gray).

flat action distribution seen in Figs. 2.14 at large time for $J < 2J_x$ is consistent with

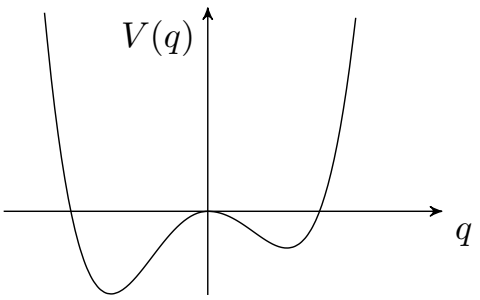
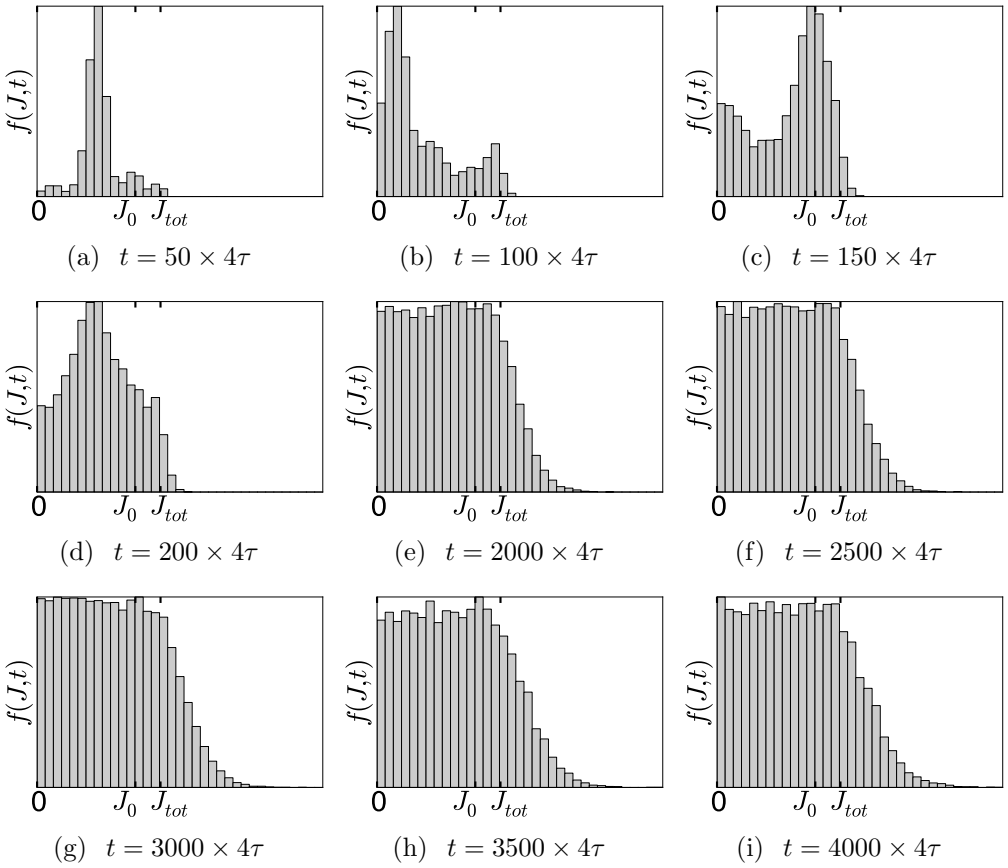


Figure 2.16: The potential at $t = 2\tau$.

the result of Refs. [6,13] that the dynamics is chaotic [13] and diffusive [6] in J in the phase space region swept by the slowly pulsating separatrix. Since the dynamics is sensitive to damping, we have used a forth order symplectic integrator [34] to avoid the friction like numerical error from the regular RK4 integrator. A two-dimensional histogram in phase space at the end of 4000th cycle is shown in Fig. 2.15. Most particles are seen to be located within the energy curve enclosing area $J = 2J_x$.

2.4.2 Repeated Cycling of the Hypothetical Curve Dynamics with Unequal Well Depths

We again consider the Hamiltonian, Eq. (2.2), but for the case where the left and right wells are of unequal depth. That is, the cycle through the square curve shown in Fig. 2.2, (λ_R, λ_L) : $(0,0) \to (1,0) \to (1,1) \to (0,1) \to (0,0)$, is replaced by a rectangular shaped cycle, (λ_R, λ_L) : $(0,0) \to (\lambda_{R0}, 0) \to (\lambda_{R0}, 1) \to (0,1) \to (0,0)$, where $\lambda_{R0} \neq 1$. Thus, at the end of the second phase of the cycle (i.e., at $t = 2\tau$) the potential is as shown in Fig. 2.16. Hence the separatrix lobes in Fig. 2.16 have different (q,p) areas: J_{xR} for the lobe in $q > 0$ and $J_{xL} > J_{xR}$ for the lobe in $q < 0$. We now consider the hypothetical curve dynamics obeyed by most points for τ large.



(a) $t = 50 \times 4\tau$ (b) $t = 100 \times 4\tau$ (c) $t = 150 \times 4\tau$

(d) $t = 200 \times 4\tau$ (e) $t = 2000 \times 4\tau$ (f) $t = 2500 \times 4\tau$

(g) $t = 3000 \times 4\tau$ (h) $t = 3500 \times 4\tau$ (i) $t = 4000 \times 4\tau$

Figure 2.17: Simulated histograms of action J after evolving the system different numbers of cycles with unequal left and right well depths are shown. We use irrational ratio between well depths $(\lambda_L, \lambda_R) = (1, (\sqrt{5}-1)/2)$. J_0 is the initial action value and $J_{tot} = J_L + J_R$.

Let J_n denotes the action at the end of the nth cycle. Then for $0 < J_0 < J_{xL} + J_{xR}$ we have that $0 < J_n < J_{xL} + J_{xR}$ for all subsequent cycles. Furthermore, following the previous arguments in Sec. 2.1.2 the hypothetical curve dynamics yields

$$J_{n+1} = \begin{cases} J_n + J_{xL} & \text{if} \quad 0 \leq J_n < J_{xR}; \\ \\ J_n - J_{xR} & \text{if} \quad J_{xR} < J_n < J_{xR} + J_{xL}. \end{cases} \tag{2.9}$$

Define $\theta_n = 2\pi J_n/(J_{xL} + J_{xR})$, $R = J_{xL}/(J_{xL} + J_{xR})$. Then Eq. (2.9) becomes

$$\theta_{n+1} = [\theta_n + 2\pi R] \ \textit{modulo}\ 2\pi. \tag{2.10}$$

Equation (2.10) can be viewed as describing successive rigid rotation of a circle by the angle $2\pi R$. If the rotation number R is rational, $R = m/l$, where m and $l < m$ are integers, then the orbit in θ is periodic with period l. E.g., for the case $J_{xL} = J_{xR} \equiv J_x$ treated in Sec. 2.4.1, $R = 1/2$, and J_n oscillates sequentially between two values separated by the amount J_x, with one value greater than J_x and one less than J_x. For R irrational, successive orbit points, θ_1, θ_2, θ_3, ... , generated by Eq. (2.10) ergodically fill the interval $0 < \theta < 2\pi$ densely and *uniformly* as $n \to \infty$. (In this case the orbit is said to be quasiperiodic [35].) Thus, starting from an initial condition $0 \leq J < J_{xL} + J_{xR}$, many cycles of the hypothetical curve evolution produce values of J that densely and uniformly fill the interval $0 < J < J_{xL} + J_{xR}$. Note that, as implied by the results in Sec. 2.4.1, in order for the hypothetical curve dynamics (as described above) to be a good representation

of the true dynamics, we require that τ be long and that n not be too large, with the limit on n increasing as τ becomes longer. In Figs. 2.17 we show numerical results of the histogram approximation of the action distribution function at times corresponding to different numbers of cycles, starting from an ensemble of 20,000 initial conditions uniformly distributed on the action curve $J_0 = 1.17$. We choose $\lambda_L = 1$ and $\lambda_R = (\sqrt{5} - 1)/2$ so that their ratio is an irrational number. As a consequence of the uniformity of the orbit density of J_n in $[0, J_L + J_R]$ for irrational R, we expect the action distribution function to approach uniformity fast in the interval $[0, J_L + J_R]$, as shown in Figs. 2.17(a-e), with a slowly growing tail in $J > J_{tot}$, as shown in Figs. 2.17(e-i).

2.5 Discussion of a Generalization to Higher Dimensionality and Chaotic Orbits

We can apply the general considerations of Sec. 2.1 to chaotic motion in higher dimensional systems by making use of the so-called Ergodic Adiabatic Invariant [36–38]. In this context, we consider Hamiltonian dynamics that is chaotic and ergodic over the energy surface. If the Hamiltonian evolves sufficiently slowly in time, at all times maintaining its chaotic and ergodic behavior, then the volume $\hat{J}$ of $(\vec{q}, \vec{p})$ space enclosed by the energy surface is an adiabatic invariant. This applies as an approximation in the limit that the time it takes a chaotic orbit to wander over the energy surface in a reasonably dense manner is much shorter than the characteristic time over which the Hamiltonian changes.

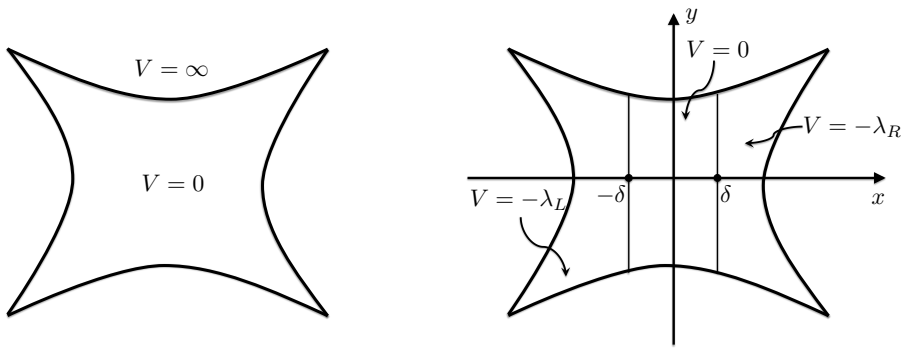


Figure 2.18: Billiard system with unchanged potential and changing potential.

As an example illustrating the Ergodic Adiabatic Invariant, consider a chaotic billiard in two spatial dimensions, as in Fig. 2.18(a) where there is a region of zero potential in the area enclosed by energy-conserving, specularly-reflecting hard walls. Due to the concave shape of the walls, the orbit of a point mass M is typically chaotic and ergodic (e.g., randomly picking some time instant far past the time at which the particle orbit was launched, the spatial probability of the orbit location is typically uniform throughout the billiard area and the probability density of travel orientation is isotropic in $(0, 2\pi)$). The volume enclosed by the energy surface is

$$\hat{J} = 2\pi M E A_B, \tag{2.11}$$

where E is the particle energy and A_B is the spatial area of the billiard. If the walls of billiard change slowly with time so that $A_B = A_B(t)$, then the particle energy will change with time $E = E(t)$ so as to keep $\hat{J}$ approximately constant. (The error in this type of adiabatic invariant is considered in Refs. [36, 37].)

To construct an example analogous to that of our one dimensional problem,

Eqs. (2.2), (2.3) and Fig. 2.3, we consider the situation shown in Fig. 2.18(b) where there are three regions of spatially constant potential, $x > \delta$ with potential,

$$V_R = -\lambda_R(t), \tag{2.12}$$

$|x| < \delta$ with potential zero at all time, and $x < -\delta$ with potential,

$$V_L = -\lambda_L(t), \tag{2.13}$$

where $\vec{\lambda}(t) = (\lambda_R(t), \lambda_L(t))$ goes through the same cycle as Fig. 2.2 and Eq. (2.3). Thus at $t = 0$ the potential is zero everywhere; in the first phase ($0 < t < \tau$) a well of negative potential and depth 1 forms in $x > \delta$; in phase 2 ($\tau < t < 2\tau$) a similar well of depth 1 forms in $x < -\delta$; in phase 3 ($2\tau < t < 3\tau$) the potential in $x > \delta$ is raised from -1 to zero thus removing the well in $x > \delta$; and in phase 4 ($3\tau < t < 4\tau$) the potential in $x < -\delta$ is raised so as to return to the situation at the beginning of the cycle (see Fig. 2.19 which shows the potential versus x on the x-axis ($y = 0$)). Letting $\hat{J}_x = 2\pi MEA_x$ where A_x is the area of the left ($x < -\delta$) or right ($x > \delta$) well, and using the same reasoning as in the one-dimensional case (Sec. 2.1.2) we again obtain

$$\hat{J}_A(4\tau) = \hat{J}_A(0) - \hat{J}_x, \tag{2.14a}$$

$$\hat{J}_B(4\tau) = \hat{J}_B(0) + \hat{J}_x, \tag{2.14b}$$

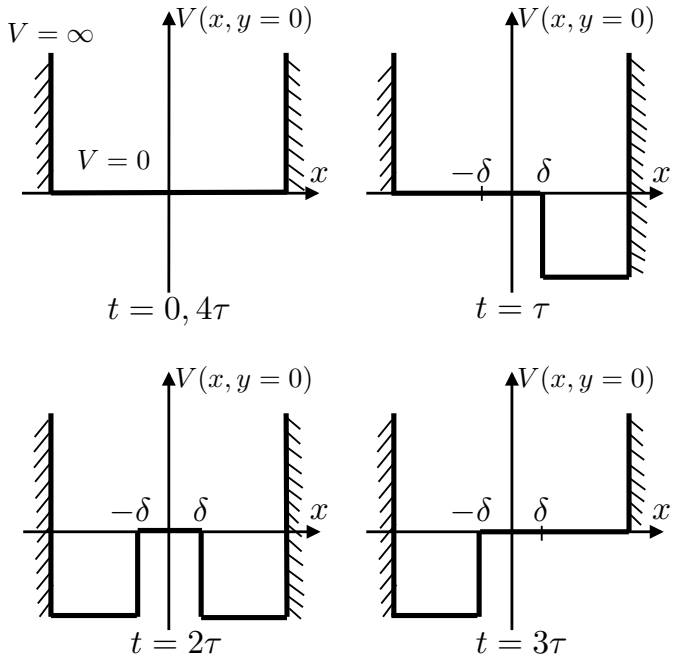


Figure 2.19: Potential function for billiard system when $t = 0$, τ, 2τ, 3τ, and 4τ.

for $2\hat{J}_x > \hat{J}_A(0) > \hat{J}_x > \hat{J}_B(0)$, which is the same as Eqs. (2.7) with the (q, p) areas J replaced by the $(\vec{q}, \vec{p})$ volumes $\hat{J}$. Thus this example displays the same type of topologically forbidden apparent interchange of energy surfaces as in the one dimensional example (Sec. 2.1).

It would be interesting to test this numerically. As opposed to the adiabatic motion associated with closed periodic orbits, the evolution in the case of ergodic adiabatic invariant has to be substantially slower (τ has to be longer) both because the chaotic ergodicity time to sample the whole energy surface can be long compared to a similar energy particle's period in a one dimensional Hamiltonian, and because in the case of the one-dimensional adiabatic invariant the error can be exponentially small in τ, while for the ergodic adiabatic invariant the error scaling is much less favorable [36, 37]. Thus numerical tests in this case might be challenging.

2.6 Conclusions

In this chapter, in order to resolve the seeming paradox discussed in the Sec. 2.1, we employ a simulation method we called "robust". The robust simulation evolves a large number of particles initially lying on an energy curve under the slowly varying Hamiltonian. New particles are added if two originally neighboring particles become significantly separated during the evolution. With this simulation technique we demonstrate the following scenario, many of the main features of which we have related to previous theoretical work [2–13, 32, 33].

Before the fixed point (separatrix) crossing the curve of points very accurately approximates an energy curve with preserved enclosed phase-space area (action). However, upon the occurrence of the fixed point (separatrix) crossing, the curve deforms and winds around the fixed point (separatrix lobe). The part of the curve that winds around the fixed point (separatrix lobe) is greatly stretched and thus can hardly be seen in the regular simulation. This effect becomes stronger as the variation of Hamiltonian is made slower. Although a regular simulation accurately generates the dynamics of individual particle, it is unable to resolve the correct topological structure for the curve dynamics.

Some additional results are that (i) a small amount of friction can cause a large change in the dynamics; (ii) consistent with Ref. [13], the effect of many repeated cycles of the protocol is to flatten the distribution of actions within the "adiabaticly allowed" action range (Sec. 2.4.1); (iii) when the two wells have unequal depths, the ideal curve dynamics become quasiperiodic (Sec. 2.4); and (iv) it appears possible

that analogous phenomena may occurs in higher dimensional system (Sec. 2.5).

Chapter 3: Resynchronization of Circadian Oscillators and the East-West Asymmetry of Jet-Lag

This chapter is based on work contained in the publication: Resynchronization of circadian oscillators and the east-west asymmetry of jet-lag, Lu, Z., Klein-Cardea, K., Lee, S., Antonsen, T. M., Girvan, M., and Ott, E., Chaos, 26(9), 094811 (2016),

3.1 Introduction

Experiments show that a region of the brain within the hypothalamus regulates the circadian rhythms of mammals [14, 15]. This region, called the suprachiasmatic nucleus (SCN), contains of the order of 10^4 pacemaker neuronal cells that exhibit oscillatory dynamics. The synchronization of these pacemaker neuronal cells is responsible for maintaining healthy circadian rhythms. Studies on SCN slice cultures from neonatal mice in which the coupling between pacemaker cells is suppressed show that the individual natural periods of SCN pacemaker cells are distributed with a mean slightly longer than 24 hours [39]. Other experiments demonstrate that the behavior of the SCN is externally affected by light induced signals transmitted from the eye and that these signals can entrain the dynamics of the SCN to a 24

hour day [39–41].

Numerous models have been formulated to study the dynamics of the SCN [42–52]. Many of these previous models rely on assumed forms for the macroscopic dynamics of the SCN [42–45], while others model the SCN as many coupled pacemaker oscillators and then numerically compute behavior of their many oscillator systems [46–52]. Here we start with a simple microscopic description of the coupled dynamics of individual pacemaker cells and then make use of an analytical reduction of the dynamics to a low dimensional, macroscopic description. Thus, our approach has the important advantage of simplicity and ease of interpretation provided by the availability of a relatively simple low dimensional macroscopic model, while retaining the realistic aspect of being based on the dynamics of many coupled oscillators. As we will show, this approach yields considerable insight, although perhaps at the expense of having to employ rather simplified modeling of the individual oscillators. Specifically, we consider a sinusoidally forced Kuramoto Model in which the forcing models diurnally periodic external inputs (e.g., light), and we employ the technique of Refs. [16, 17] to extract a macroscopic description from the microscopic model. We are specifically interested in using this model to study how human circadian rhythms are affected by cross-time-zone travel (i.e., the phenomenon of jet-lag) and how SCN pacemaker cells become re-entrained to a new time-zone.

3.2 Model

We model the SCN using $N \gg 1$ globally coupled phase oscillators with individual phase angles θ_i $(i = 1, ..., N)$. The coupling strength between pacemaker cells is assumed to be uniform for all connections and denoted by K. We model the 24 hour period external light source by introducing a temporally sinusoidal external driving term. The evolution for each oscillator is given by

$$\frac{d\theta_i}{dt} = \omega_i + \frac{K}{N}\sum_{i=1}^{N}\sin(\theta_j - \theta_i) + F\sin(\sigma t - \theta_i + p(t)), \tag{3.1}$$

where ω_i is the natural frequency for oscillator i when it is decoupled from other oscillators, $\sigma = 2\pi/24$ $(\text{hrs})^{-1}$ is the day-night frequency of the external drive and F is its amplitude. The phase quantity p is determined by the time-zone in which the individual is located and takes values between $-\pi$ and π. The variables satisfying Eq. (3.1) may be interpreted as follows. The oscillator phases advance in time with "fully awake (noon)", for example, corresponding to the set of times $\theta_i(t) = 2\pi n$ with n integer. If all oscillators have a 24-hour period, $\omega_i = \sigma$, Eq. (3.1) indicates that each phase is attracted to the same value, $\theta_i = \sigma t + p(t)$. If time t is defined to be the Greenwich time, then $p = 0$ corresponds to being in Greenwich. Local noon to the east of Greenwich occurs earlier than noon in Greenwich. Thus, $p > 0$ to the east and $p < 0$ to the west of Greenwich. During travel p changes with time. For simplicity, in our study we assume that all travel is very fast and approximate it as instantaneous so that we will take $p(t)$ to change discontinuously from one constant

value to another where the time of discontinuity corresponds to the time of travel. See Refs. [53, 54] for previous work on the dynamics of Eq. (3.1) (not specifically in the context of jet-lag, i.e., with $p(t) = 0$).

With our fast travel assumption,

$$p(t) = \begin{cases} p_1, & \text{if } t \leq \tau \\ p_2, & \text{if } t > \tau, \end{cases} \tag{3.2}$$

where the fast travel takes place at time τ and the travel starts from a time-zone with phase p_1 and ends in another time-zone with phase p_2. We further suppose that the system has been in the $p = p_1$ state for a long time before the travel takes place, so that, by the time $t = \tau$, the system has evolved to its $p = p_1$ attractor. Thus we model recovery from jet-lag by following the solution of Eq. (3.1) for $t \geq \tau$ with $p = p_2$ from an initial condition on the attractor of Eq. (3.1) with $p = p_1$.

We take the oscillators' natural frequency distribution $g(\omega)$ to be Lorentzian,

$$g(\omega) = \frac{\Delta}{\pi \left[(\omega - \omega_0)^2 + \Delta^2\right]}, \tag{3.3}$$

where Δ and ω_0 denote the spread and the center of the natural frequency distribution, respectively. Similar to previous studies on forced Kuramoto models [16, 54], we reduce the $N \gg 1$ dimensional dynamics of the system to the dynamics of a

complex order parameter z in a rotating frame with frequency σ,

$$z = \frac{1}{N}\sum_{j=1}^{N} e^{i[\theta_j(t)-\sigma t-p_2]}. \tag{3.4}$$

The dynamics of this order parameter for $N \gg 1$, and $t \geq \tau$ is governed by the equation [54],

$$\dot{z} = \frac{1}{2}[(Kz+F) - z^2(Kz+F)^*] - (\Delta + i\Omega)z, \tag{3.5}$$

where $\Omega = \sigma - \omega_0$, $\dot{z} = dz(t)/dt$, and $\Omega = 0$ corresponds to the case where the center of the natural frequencies of the oscillators are the same as the external frequency σ. Since many experimental studies show longer-than-24-hours-free-running-period of human circadian rhythms [55–60], the average frequency of SCN pacemaker cells in typical humans is thought to have a period longer than 24 hours, and thus Ω is positive.

If an individual is in a steady entrained state before travel, and at time $t = \tau$ travels from a time-zone with $p = p_1$ to a timezone with $p = p_2$, then, upon his/her arrival, the individual's order parameter is $z_{\rm st}e^{i(p_1-p_2)}$, where $z_{\rm st}$ is the entrained state in the new time-zone, i.e., $z_{\rm st}$ is the stable fixed point of Eq. (3.5), assuming that it exists. Note that Eq. (3.5) and $z_{\rm st}$ do not depend on the phase of the destination time-zone, p_2.

Because an individual in the new time-zone experiences misalignment between his/her biological time and the local time, recovery from cross-time-zone travel takes

place in the following days as the order parameter gradually evolves back to the stable fixed point $z_{\rm st}$. That is, recovery from jet-lag is described by Eq. (3.5) for $t \geq \tau$ with the initial condition

$$z(\tau) = z_{\rm st} e^{i\Delta p}, \ \Delta p = p_1 - p_2, \tag{3.6}$$

where $\Delta p > 0$ for westward travel, $\Delta p < 0$ for eastward travel, and traveling to a time-zone with a one hour difference in clock setting corresponds to $|\Delta p| = 2\pi/24$. During recovery, in what follows, we will quantify the level of the deviation from full recovery by the distance between the order parameter and the stable fixed point, $|z(t) - z_{\rm st}|$.

The model we have studied is highly simplified, but, as previously noted, offers the advantage of having a individual oscillator basis and reducible to a macroscopic description. Some aspects that are not included but have experimental basis are spatially structured coupling within the SCN [39], both attractive and repulsive interactions of SCN pacemaker cells [49], and potential influence on diurnal timing from other organs of the body [61].

3.3 Dynamics

The attractors of Eq. (3.5) have been studied in Refs. [53, 54]. It was shown that, with different sets of the parameters, K, F, Δ, and Ω, the system exhibits different dynamics separated by bifurcations.

Figures 3.1(a-c) show dynamical phase flow portraits for the three main types

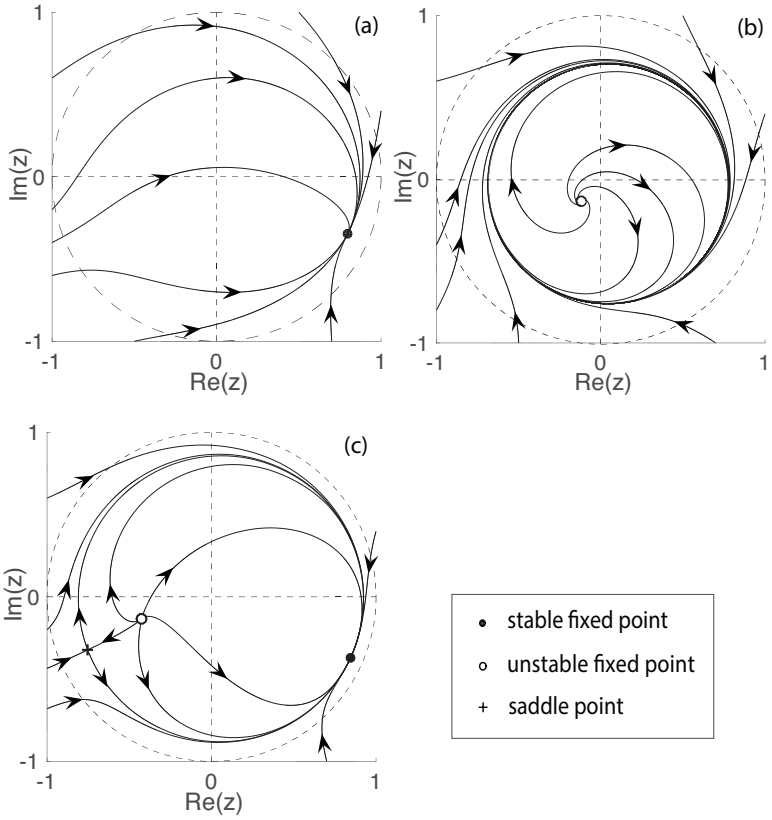


Figure 3.1: These three figures show phase portraits of the order parameter z for three types of dynamics. Stable fixed points, unstable fixed points and saddle points are denoted by solid dots, empty circles, and crosses, respectively. Panel (a) shows the Type A dynamics, where there is one and only one fixed point, and it is stable. Panel (b) shows the Type B dynamics with one unstable fixed point and one stable limit cycle. Panel (c) shows the Type C dynamics with one stable, one unstable, and one saddle fixed point. The parameters for (a), (b), and (c) are ($K = 4.5\Delta$, $F = 3.5\Delta$, $\Omega = 1.4\Delta$), ($K = 4.5\Delta$, $F = 0.65\Delta$, $\Omega = 1.4\Delta$) and ($K = 10.0\Delta$, $F = 3.5\Delta$, $\Omega = 1.4\Delta$), respectively.

of dynamics for Eq. (3.5). Following Antonsen *et al.* [53], we label these three types Type A, Type B, and Type C, where each type occurs in its corresponding region of $(K/\Delta, F/\Delta, \Omega/\Delta)$ parameter space, and the three panels of Fig. 3.1 are flows computed at specific parameters sets in the appropriate region.

For Type A dynamics (Fig. 3.1(a)), the system has one fixed point that is stable. For Type B dynamics (Fig. 3.1(b)), there is a limit cycle that is stable and an unstable fixed point enclosed within the limit cycle. For Type C dynamics (Fig. 3.1(c)) there are three fixed points, one stable, one unstable and one of saddle type, and the unstable manifold of the saddle flows to the stable fixed point.

By definition (3.4) the magnitude of the order parameter satisfies $|z| \leq 1$; thus, we are only interested in the dynamics within the circle $|z| = 1$ which is shown as the dashed circles in Figs. 3.1(a,b,c). Also from Eq. (3.4), larger $|z|$ corresponds to greater synchrony, and $|z| = 1$ corresponds to perfect synchrony (i.e., all the θ_j are equal).

Note from Figs. 3.1(a) and 3.1(c) that the stable fixed point lies below the positive real axis. This reflects the parameter choice $\Omega > 0$, corresponding to the typical situations for humans (i.e., a slower SCN frequency, represented by ω_0, as compared to the 24 hour driving, represented by the frequency σ). Note that this negative phase of z_{st} corresponds to a human circadian response whose oscillation phase lags the forcing oscillation phase.

In Type B dynamics the attractor is a limit cycle and thus not fully entrained to the forcing. This case would correspond to an individual whose awake/sleep times naturally vary from day to day, even if the driving (e.g., sunlight) is the same every

day. We note, however, that if Type B dynamics applies with parameter values near the bifurcation point to Type A dynamics, then the limit cycle is a relatively small loop around the unstable fixed point that was stable before the bifurcation. In this case there may be little practical difference between Type B dynamics and a case with a stable fixed point.

In our subsequent parameter studies, for simplicity, we will investigate parameter regions where the dynamics is of either Type A or Type C. Thus Type B dynamics will not be further considered in this chapter.

3.3.1 Recovery from Jet-Lag

For the same number of time-zones crossed, several studies have indicated that recovery from jet-lag is typically faster following westward travel as compared with eastward travel [18, 62–65]. A primary goal of our modeling efforts is to understand the underlying reason for this asymmetry.

Naively, we might expect that when an individual travels eastward crossing less than 12 time-zones, his/her intrinsic phase is delayed compared to the local time phase of the destination, and thus the individual re-entrains his/her intrinsic clock by phase advancing. Similarly, we might expect that an individual who travels across less than 12 time-zones westward should re-entrain his/her intrinsic clock by phase delaying. However, counter intuitively, 'wrong direction' re-entrainment has been found for large enough eastward travels [44, 66, 67], where, for example, an individual who travels eastward crossing 8-10 timezones re-entrains his/her intrinsic

clock by phase-delaying, instead of phase advancing.

In the following we discuss this temporal and directional asymmetry in jet-lag recovery using the dynamics of our model. We first illustrate Type C dynamics in Fig. 3.2(a) in which we show order parameter trajectories resulting from trips crossing different numbers of time-zones. We plot four trajectories corresponding to

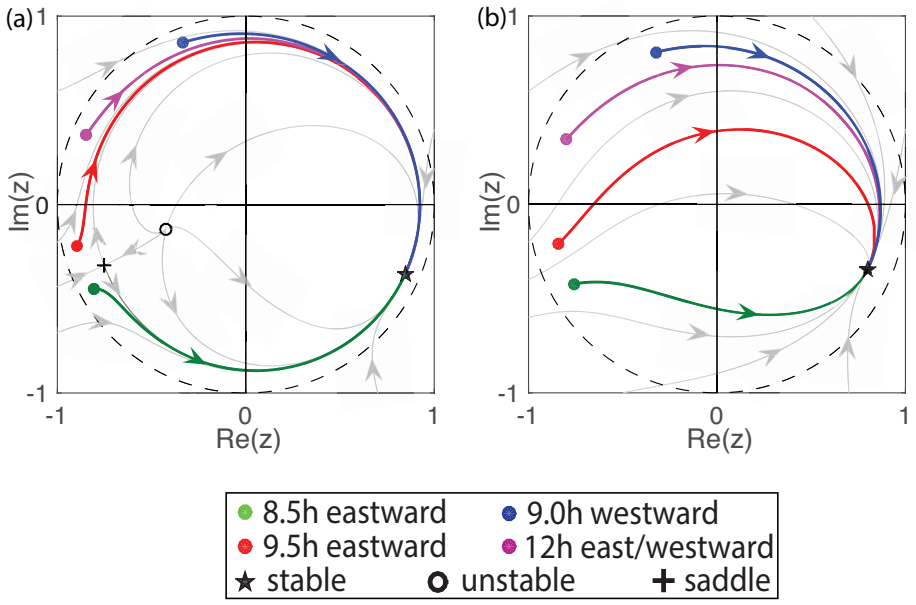


Figure 3.2: The trajectories of the order parameters corresponding to different time-zone travels for Type C and A dynamics (panels (a) and (b) respectively). The parameters for (a) and (b) are ($K = 10\Delta$, $F = 3.5\Delta$, $\Omega = 1.4\Delta$) and ($K = 4.5\Delta$, $F = 3.5\Delta$, $\Omega = 1.4\Delta$), respectively.

jet-lags from traveling across 9 and 12 hours westward and across 8.5 and 9.5 hours eastward for the parameters given in the figure caption, where the initial condition for each trip is determined by using Eq. (3.6). (In Subsections 3.3.2 and 3.3.3, we discuss exploration of the parameter space.) The four trajectories in Fig. 3.2(a) are respectively plotted as blue, cyan, green and red curves superposed on a copy of the phase space plot of Fig. 3.1(c) (plotted in grey). For the westward 9 hour trip,

the order parameter (in blue) evolves clockwise, decreasing its phase so as to realign to the stable fixed point. For the eastward 8.5 hour trip, the order parameter (in green) evolves counterclockwise approaching the stable fixed point. In contrast, for the eastward 9.5 hour trip, the order parameter moves clockwise. This asymmetry results from the presence of the saddle point. Specifically, although the eastward 8.5 and 9.5 hour trips have close initial starting points in the order parameter space, they lie on opposite sides of the stable manifold of the saddle point, and thus they flow in two different directions. The crossover point between delays and advances of the phase lies on the stable manifolds. An eastward journey with a time-zone shift pass the crossover point results in phase-delaying recovery, while an eastward journey with a time-zone shift smaller than the crosser-over point results in phase-advancing recovery, as would typically be expected of eastward travel. We note that according to our model, an initial condition exactly on the stable manifold of the saddle would never flow to the stable fixed point and initial conditions close to the stable manifold could take a very long time to approach the vicinity of the stable fixed point (meaning recovery from jet-lag would correspondingly be very drawn-out). In a real situation, however, we do not expect arbitrarily long recovery times due to factors omitted from the model, such as day-to-day deviations in the external drive term due, for example, to light fluctuations resulting from varying weather patterns.

Analogous to Fig. 3.2(a) (which applies for parameters where there is Type C dynamics), Fig. 3.2(b) shows order parameter trajectories for a set of parameters for which Type A dynamics applies. In going from Fig. 3.2(a) to Fig. 3.2(b), the repeller

fixed point and the saddle fixed point, along with the stable and unstable manifolds of the saddle, have annihilated. Nonetheless, the order parameter trajectories for Fig. 3.2(a) and Fig. 3.2(b) are similar. This is because of the local character of the saddle-node bifurcation and because the parameter values of Fig. 3.2(b) and Fig. 3.2(a) are not too far from the saddle-node bifurcation point. Note that due to the absence of the saddle, Type A dynamics of the model cannot lead to arbitrarily long jet-lag recovery times (in contrast with Type C dynamics).

3.3.2 Reference Parameter Set

We now attempt to estimate a reasonable set of the parameters (K, F, Ω and Δ) that might be thought to correspond to a typical healthy person. However, we recognize that very substantial variation of these parameters is indicated by the observation that the severity of jet-lag has large variation from one individual to another. Thus we will use our putative 'typical' parameter set as a reference point from which we will study the effect of parameter deviations. Our choice of a typical parameter set is based on the following:

1. Pacemaker cells in the SCN should be in a fairly synchronized state, which suggests that the stable fixed point of the order parameter has a fairly large amplitude (i.e., $|z| < 1$ should not be small compared with one).

2. Empirical studies (Refs. [56, 68]) suggest that an individual confined to a dark enclosure may have substantial (but noisy) regularity in his sleep-awake cycle. Thus we surmise that the system can synchronize when $F = 0$.

3. No direct measurements of SCN oscillators are available for human, and thus any estimation of the oscillator distribution must be based on indirect deductions. Observations of blind people [55] and of an individual in a dark enclosure [56], as well as other earlier experimental results [57, 58], suggested that typical circadian rhythms have a period very roughly in the range 24.5-26 hours when there is no external drive. However, it has been argued [59] that such observations are not indicative of the natural period of the relevant biological oscillators, and, using a different methodology, Ref. [59] has estimated an average period of roughly 24.2 hours. However, another, more recent study [60], estimated a period of 24.5 hours. Here, we will take 24.5 hours as value for the average period. [1]

4. The speed for an individual to recover from jet-lag is different for eastward and westward travels [18]. For example, it is sometimes claimed that, on average, an individual approximately retrains his/her phase by 1 hour per day after an eastward travel and retrains his/her phase by 1.5 hour per day after a westward travel. Also, 9-hour eastward travel typically corresponds to the border between jet-lag recovery by phase advancing or delaying [69, 70], and thus is considered most disruptive.

Based on the above and on numerical simulations of our model (to incorporate

[1] We also explored cases with an average oscillator period of 24.2 hours consistent with Ref. [59]. However, with this smaller period it was not possible to find a set of the other parameters of the model system (K, F, Δ) that simultaneously yielded results approximately consistent with empirical studies. If the 24.2 hour results of Ref. [59] actually applies, the above may indicate that our model is too simple to be trusted for quantitative prediction.

the information from point 4), we choose

$$\Delta = 3.8 \times 10^{-3} \ (\text{rad} \cdot \text{hour}^{-1}), \tag{3.7}$$

$$\Omega = 1.4\Delta = 2\pi/24 - 2\pi/24.5, \tag{3.8}$$

$$K = 4.5\Delta, \tag{3.9}$$

$$F = 3.5\Delta, \tag{3.10}$$

as our reference parameter set. (For this set Type A dynamics applies.) The corresponding recovery curves from several travels eastward and westward are shown in Fig. 3.3, where the vertical axis is $|z(t) - z_{\text{st}}|$, which we regard as a quantitative measure of the severity of jet-lag at time t. We take recovery from jet-lag to have occurred when $|z(t) - z_{\text{st}}|$ drops below 0.2 (the horizontal black line). We see that for this parameter set there is a strong asymmetry in recovery, with westward recovery being significantly faster, and that the longest recovery time occurs for 9 hour time-zone eastward travel.

3.3.3 Parameter Dependence

As previously mentioned, the effects of cross-time-zone travel vary substantially from individual to individual. With this motivation, we investigate the parameter dependence of the model dynamics and its implications for recovery from jet-lag.

Dividing Eq. (3.5) by Δ, we see that keeping K/Δ, F/Δ and Ω/Δ constant

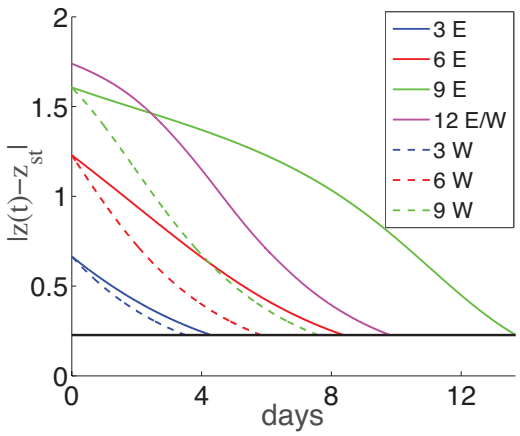


Figure 3.3: Recovery curve from several eastward and westward travels with different numbers of time-zones crossed. The parameters for this simulation are from Eqs. (3.7)-(3.10). The vertical axis is the distance between the order parameter and the stable fixed point $|z(t) - z_{\rm st}|$. The dashed lines are recovery curves for westward travels and the solid lines are for eastward travels. We define recovery from jet-lag to have occurred when $|z(t) - z_{\rm st}|$ drops below 0.2 (the horizontal black line).

while varying the value of Δ is equivalent to changing the time scale of the dynamics. Thus, in the following, we keep Δ constant at our reference value, Eq. (3.7), and study how the dynamics change when K, F, and Ω are individually varied away from the reference parameter set (Eqs. (3.8)-(3.10)).

The parameter K, which represents the oscillator coupling strength, controls the degree of synchronization for fixing values of the external drive, F. In Fig. 3.4(a) we show the recovery time in a range of K for several cross-time-zone journeys with the other two parameters at their reference values ($F = 3.5\Delta$ and $\Omega = 1.4\Delta$). For K/Δ between 2.0 and about 8.0, there is only one fixed point that is stable (Type A dynamics). The magnitude of the stable fixed point increases as K increases (not shown), corresponding to increasing strength of the degree of synchronization. When K exceeds $K \approx 8.0\Delta$, two other fixed points, one unstable and one saddle, appear through a saddle-node bifurcation (Type C dynamics). The large recovery time for

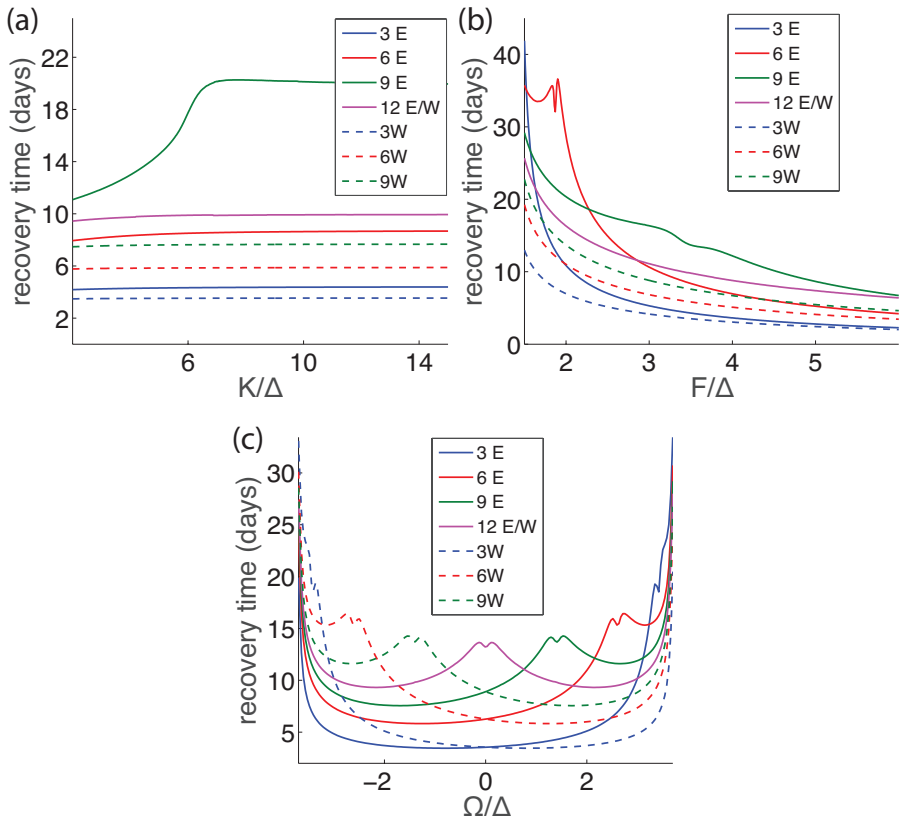


Figure 3.4: Recovery time dependence on K (panel (a)), F (panel (b)), and Ω (panel (c)) with the nonvarying parameters kept at the reference values in Eqs. (3.7)-(3.10).

eastward 9-hour travel is due to the newly born saddle point (see Figs. 3.1 and 3.2) that is very close to the the eastward 9-hour travel initial point. In simulations with other parameter sets (not shown), a sharp, narrow spike in the recovery time can occur in plots like those in Fig. 3.4(a) when, as K/Δ is varied, the stable manifold of the saddle point crosses the initial point of the jet-lag simulation, causing the recovery time to go to infinity.

The parameter F controls the magnitude of the external drive which is meant to capture the effects of factors such as the sensitivity of an individual's response to sunlight, cloudiness, geographical latitude, and seasonality. In Fig. 3.4(b) we show the recovery time in a range of F for several cross-time-zone journeys with the other two parameters at their reference values ($K = 4.5\Delta$ and $\Omega = 1.4\Delta$). In the parameter range illustrated, the dynamics are always Type A. Not surprisingly, the recovery times for all seven cross-time-zone journeys decrease with increasing F. For increasingly large values of F, the stable fixed point approaches the positive real z-axis (not shown), reflecting decrease in the difference between the external drive phase and the internal overall phase. The decrease of recovery time with increasing F is consistent with the idea that recovery from jet-lag is facilitated by being outdoors in bright daytime sunlight (effectively increasing F).

The parameter $\Omega = \sigma - \omega_0$ is the difference between the external drive frequency σ (corresponding to a 24 hour day) and the average of the pacemaker cells' natural frequencies ω_0. In Fig. 3.4(c) we show the recovery time in a range of Ω from -3.7Δ to 3.7Δ for several cross-time-zone journeys with the other two parameters at their reference values ($K = 4.5\Delta$ and $F = 3.5\Delta$). From Fig. 3.4(c) we notice

that the eastward and the westward recovery time curves (dashed and solid curves of the same color) are reflections of each other about the $\Omega/\Delta = 0$ axis. This can be seen from the fact that the complex conjugate of Eq. (3.5) with Ω replaced by $-\Omega$ yields an equation for $z^*(t)$ that is the same as the original equation (Eq. (3.5)) for $z(t)$.

3.4 Conclusions

In this chapter we present a model of circadian dynamics that is based on the microscopic dynamics of individual pacemaker cells in the suprachiasmatic nucleus (SCN). Our model employs an extremely simple phase oscillator description of the individual SCN cells. This description, while not reflecting the true complexity of the mechanisms governing the individual cells (see, e.g., Ref. [46, 47]), has the advantage that in the large N limit (N = the number of pacemaker cells), it allows an exact reduction of the high dimensional microscopic dynamics to that of a relatively simple low dimensional system for the evolution of the macroscopic system state. Using this low dimensional macroscopic description, we model the dynamics of resynchronization of the SCN following rapid cross-time-zone travel, i.e., jet-lag. We believe that this approach has resulted in increased understanding of the East/West jet-lag asymmetry and of the origin of potential differences in the responses of different individuals to cross-time-zone travel.

While our approach provides insights into the key features of jet-lag from a dynamical systems perspective, a number of avenues remain for further exploration.

For example, our basic uncoupled phase oscillator ($d\theta_i/dt = \omega_i$) could be replaced by other, potentially more realistic, phase oscillator descriptions [71, 72], while still allowing a reduction to low dimensional dynamics. Also, because many features concerning the coupling of SCN pacemaker cells are unknown, as well as for the sake of simplicity, we have assumed an all-to-all network topology for the connection between SCN pacemaker cells. We note, however, that experience with other networked systems suggests that heterogeneous connectivity patterns may play an important role in the resynchronization of circadian oscillators in individuals experiencing jet-lag. Finally, we have implemented rapid cross-time-zone travel as a discontinuous phase shift in a sinusoidal external drive term, but more complex external drive functions may be appropriate and could even be used, for example, to explore the effectiveness of different strategies for combating jet-lag.

Chapter 4: Inhibitory Neurons Promote Robust Critical Firing Dynamics in Networks of Integrate-and-Fire Neurons

This chapter is based on work contained in the publication: Inhibitory neurons promote robust critical firing dynamics in networks of integrate-and-fire neurons, Lu, Z., Squires, S., Ott, E., and Girvan, M., Phys Rev E, 94(6), 062309 (2016),

4.1 Introduction

Experimental studies [19–27] show that neural networks in vivo and in vitro exhibit avalanches of firings that obey power law scaling. For example, researchers have observed neural avalanches with power-law distributed sizes in multi-electrode recordings of neuronal cultures and slices of rat cortex [19, 20]. In addition, various models have been used to explain this phenomenon [73–79]. Furthermore, it has been argued, with experimental support, that this type of dynamics is crucial for information processing in the brain (e.g., see [80] and references therein). See [81] for a critical review.

In Ref. [73] Eurich and Herrmann studied a continuous-state/discrete-time model of globally coupled threshold firing neurons, which they found produced

power-law avalanche size distributions consistent with an exponent of $-3/2$. In this model, 'energy' is supplied to randomly chosen neurons at each time step, and, when the 'energy' of a neuron exceeds a threshold, it immediately fires. The total 'energy' of firing neurons is then evenly distributed to all neurons in the system, with or without net dissipation.

Levina et al. [74, 75] proposed similar all-to-all network models, but with biologically more realistic dynamical synapses. In these models, the signals emitted by firing neurons are transmitted through synapses, and the transmission ability of a synapse decreases when its presynaptic neuron fires more frequently (this models the consumption of firing resources within the synapse). In a certain parameter regime, which they showed was remarkably large, they found that the behavior was 'critical' in that there was an approximate power law distribution of firing avalanche size with exponent $-3/2$[1]. The main point of their papers was the robustness of the critical behavior, attributed to the introduced dynamics of synapse strength dependence on firing frequency. We emphasize that in this chapter we follow Refs. [74, 75] and adopt the heuristic definition of the word 'critical' to refer to a parameter range in which the expected power law avalanches statistics approximately applies.

Firing avalanches have also been studied in models employing binary-state/discrete-time neurons on random networks [76, 77]. In these models, any neuron can be either in a firing or quiescent state. On each time step each neuron fires with a probability depending only on the sum of the incoming signals it received on the previous time

[1] We note, however, that other authors have considered more sophisticated criteria which they use to test the existence of 'criticality' [82], and that the latter work has recently prompted critical discussion [83]

step. In contrast to the models in Refs. [73–75], the propensity to fire does not build up due to signals received at earlier times. In the model studied in Ref. [76], all neurons are excitatory, i.e., a firing neuron increases the next step firing probabilities of the neurons to which it sends signals. Larremore et al. obtained critical power-law distributions of avalanche size and duration with exponents $-3/2$ and -2, respectively, by tuning the largest eigenvalue of the network adjacency matrix to the critical value of one. In Ref. [77], Larremore et al. extended the earlier model by introducing inhibitory neurons. In contrast to excitatory neurons, upon firing, an inhibitory neuron sends out signals that decrease the firing probabilities of its connected post-synaptic neurons. Larremore et al. showed that the effect of introducing inhibitory neurons is to make the critical firing behavior 'ceaseless', i.e., starting from a completely quiescent state, a single externally inducing firing event can trigger an endless sequence of firing activity. Within this ceaseless activity, they observed avalanche dynamics in which the number of firing neurons fluctuating wildly from very low to very high without entering a state where every single neuron is quiescent.

In this chapter, we propose a discrete-time/discrete-state firing neuron network model with both excitatory and inhibitory neurons. However, different from the model in [77], each neuron undergoes an integrate-and-fire type process, in which the probability of firing at time t depends, not only on the signals received at time $t-1$, but on all signals received since it last fired. Details of the model are presented in Sec. 4.2. In Sec. 4.3, we show numerically that this model exhibits non-ceaseless firing with avalanches in the sub-critical region, ceaseless firing in the super-critical

region, and critical firing behavior with power-law distributions of avalanche size and duration with exponents $-3/2$ and -2. In Sec. 4.4, using a mean-field type analysis, the condition for criticality and the accompanying dynamics are analyzed. Simulation results and comparison with the theory in Sec. 4.4 are given in Sec. 4.5. In Sec. 4.6, we report the interesting finding that, the inclusion of inhibitory neurons promotes robustness to parameter perturbations of the critical power law regime. Concluding results and discussion are presented in Sec. 4.7. Overall, the main result of this chapter is that the simultaneous presence of *both* a significant fraction of inhibitory neurons *and* integrate-and-fire dynamics promotes robust criticality over a greatly enhanced range of system parameters.

4.2 Discrete-State/Discrete-Time Neural Network Model

The state of a neuron is represented by an integer 'potential' value $x_i(t) \in \{0, 1, 2, ..., M\}$ where $i \in \{0, 1, 2, ..., N\}$ and, $t \in \mathbb{N}$ is the discretized time. On each time step, the states for all neurons are updated simultaneously. Specifically, when the network is active, the potential value for any neuron at time $t+1$ is determined by its potential value $x_i(t)$ at t, and the net signal $s_i(t)$ it receives from its fired excitatory and inhibitory neighbors. If the combination of the net signal received by a neuron and the neuron's current state value exceeds a threshold M, the neuron fires and its state is reset to zero. We define the firing function $f_i(t)$ for neuron i such that $f_i(t) = 1$ if neuron i fires at time t, and $f_i(t) = 0$ otherwise.

We allow each neuron to be excitatory or inhibitory and we denote the fraction

of the neurons that are excitatory by α. The network topology is defined by two sparse adjacency matrices, A^+ and A^-, where A^+ contains the outgoing links from excitatory neurons, and A^- contains the outgoing links from inhibitory neurons. Specifically, $A^{\pm}_{ij} = 1$ if there is an outgoing connection from neuron j to neuron i and $A^{\pm}_{ij} = 0$ otherwise. Note that we designate each neuron as strictly excitatory or inhibitory. That is, if a neuron i has at least one outgoing excitatory (inhibitory) link, i.e., $A^+_{ij} = 1$ for some j ($A^-_{ij} = 1$ for some j), then it can have no outgoing inhibitory (excitatory) links, i.e., $A^-_{ik} = 0$ for all k ($A^+_{ik} = 0$ for all k).

In terms of the quantities A^+, A^- and f_i, the net signal to neuron i from other neurons is

$$s_i(t) = \sum_j (A^+_{ij}(t) - A^-_{ij}(t)) f_j(t). \tag{4.1}$$

In addition, if the system is quiescent at time t (i.e., $f_j(t) = 0$ for all j), then we randomly choose one of the N neurons r and give it an input of size 1. We represent this input by

$$d_i(t+1) = \begin{cases} \delta_{i,r} & \text{if } f_j(t) = 0 \text{ for all } j, \\ 0 & \text{otherwise,} \end{cases} \tag{4.2}$$

where $\delta_{i,r}$ in Eq. (4.2) is the Kronecker delta. Our state-updating protocol is now described by Eqs. (4.3)-(4.5).

$$\tilde{x}_i(t+1) = x_i(t) + s_i(t) + d_i(t), \tag{4.3}$$

$$f_i(t+1) = \begin{cases} 1 & \text{if } \tilde{x}_i(t+1) > M, \\ 0 & \text{otherwise,} \end{cases} \tag{4.4}$$

$$x_i(t+1) = \begin{cases} 0 & \text{if } \tilde{x}_i(t+1) < 0, \\ 0 & \text{if } \tilde{x}_i(t+1) > M, \\ \tilde{x}_i(t+1) & \text{otherwise .} \end{cases} \tag{4.5}$$

The quantity $\tilde{x}_i(t)$ defined in Eq. (4.3) is an intermediate state value which determines whether neuron i fires, as determined by Eq. (4.4), where M is the threshold state value of $\tilde{x}_i$ above which neuron i fires. For $M = 0$, the probability that a neuron fires depends only on firing events of its inputs in the previous time step, as in Ref. [77]. For $M > 0$, we have integrate-and-fire type behavior, and the role of 'integration,' i.e. keeping a running sum of incoming signals between firing events, is enhanced as M is increased.

In our study, we tested three types of directed sparse random Erdős-Rényi network setups, 'annealed', 'partially-annealed', and 'quenched'. For annealed random networks, at each time t, new matrices $A^+(t)$ and $A^-(t)$ are randomly generated before applying Eq. (4.1). For the partially-annealed case, at the beginning of the simulation, two matrices $\tilde{A}^+$ and $\tilde{A}^-$ are generated with twice as many links as specified by the average degree k. Then, at each time in the simulation, the matrices $A^+(t)$ and $A^-(t)$ are constructed by randomly selecting half of the non-zero elements

from $\tilde{A}^+$ and $\tilde{A}^-$. For the quenched case, A^+ and A^- are initially randomly generated and fixed during the simulation. (We have introduced the partially-annealed protocol since, for the purpose of comparison, we find it useful to have a protocol 'mid-way' between the quenched and annealed protocols.)

4.3 Avalanches and Ceaseless Dynamics

4.3.1 Firing Avalanches

In our numerical simulations, depending on parameters, we find that the dynamics can be either ceaseless, or else an initial firing state can eventually evolve with time to a quiescent state. In the latter case, as specified by Eq. (4.2), when

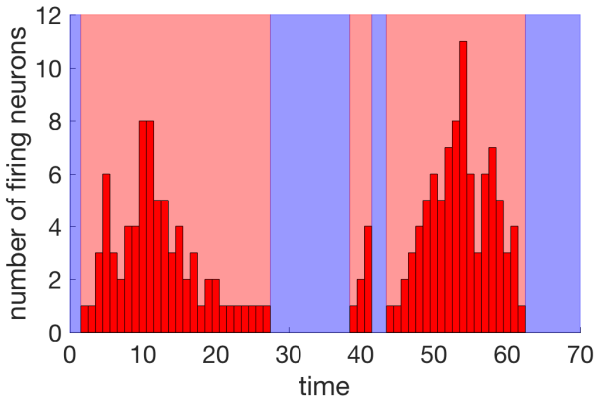


Figure 4.1: Three avalanches from a simulation of an annealed random network with $N = 10000$, $M = 5$, $\alpha = 0.9$, and $k = 6$. The avalanches are marked by the red shaded region while the quiescent regions are shaded blue.

quiescence occurs, we randomly choose one neuron and increase its potential by 1. For this type of dynamics, we define an 'avalanche' as starting from the first firing event after the onset of quiescence and ending when firing dies out. The size S of the avalanche is defined as the total number of firing events within the avalanche

and the duration T of the avalanche is defined as the time span for the avalanche.

Figures 4.1 and 4.2 illustrate avalanche dynamics for the case in which the fraction of network neurons that are excitatory is $\alpha = 0.9$, the firing threshold is $M = 5$, and different values of the average degree k. In Fig. 4.1 we show examples of three avalanches (in red) spaced by quiescent periods (in blue). Due to the external drive, at the beginning of any avalanche there is 1 neuron in the firing state. We see that, although the external drive is absent during avalanches, a single firing event induced by the external drive can trigger a sequence of more and more firings until, at some time point, the firings cease.

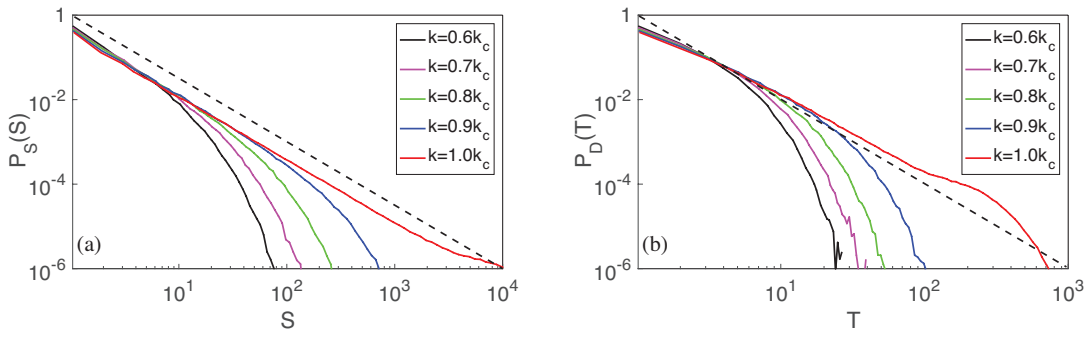


Figure 4.2: Histograms for (a) the avalanche size S, and (b) the avalanche duration T. Simulation parameters are, $N = 10000$, $M = 5$, $\alpha = 0.9$. The dashed line indicate the critical power $-3/2$ for size and -2 for duration. The k_c denotes the network degree where critical power-law scaling avalanches are predicted (Sec. 4.4).

The size and duration of avalanches are important quantities for characterizing the firing dynamics. In particular, in other neuron firing models it has been found [76], that for appropriately tuned parameter sets, histograms of avalanche size $P_S(S)$ and duration $P_D(T)$ from long simulations show critical power-law distributions, with critical exponents $-3/2$ and -2, and the values of these exponents have been derived. In our model, critical power law scaling avalanches are observed at the

critical network degree, denoted by k_c. (The dependence of k_c on model parameters, M and α, is derived in Sec. 4.4 and given by Eq. 4.23.) In Figs. 4.2(a) and 4.2(b), we show histograms of avalanche size S and duration T from simulations with $N = 10000$ neurons on annealed random networks. As shown in Figs. 4.2(a) and 4.2(b), as k approaches the critical degree k_c from below, the histograms become more and more power-law. At $k = k_c$ (red curves), power-law scaling with the expected exponents ($-3/2$ and -2, respectively) becomes evident over many decades for both S and T.

Due to the finite size N of the network, the power-law distributions eventually get cutoff if S or T is too large. In particular, even in the critical case, we expect deviations from the large N limit when the size of an avalanche S becomes larger than the size of the system $N = 10^4$. Thus we only plot $P_S(S)$ for $S \leq 10^4$. The agreement of $P_S(S)$ at $k = k_c$ (red curve in Fig. 4.2 (a)) with a power law with $-3/2$ exponent is seen to be quite good, extend over ~ 6 decades in P_S (corresponding to ~ 4 decades in S). On the other hand, referring to Fig. 4.2 (b), the agreement of $P_D(T)$ with the predicted -2 exponent is not as good and extends over fewer decades. Since the $P_S(S)$ plot seems to be a better indicator of criticality, for our investigation of the robustness of criticality in Sec. 4.6 we concentrate on $P_S(S)$.

In Figs. 4.2(a) and 4.2(b) we only show results for $k/k_c \leq 1$, because, as soon as k/k_c becomes slightly larger than 1, the dynamics becomes ceaseless and avalanches do not occur. We emphasize, however, that Figs. 4.2(a) and 4.2(b) are for $\alpha = 0.9$, and that (Sec. 4.3.D and 4.6) at lower α (i.e., when the fraction of neurons that are inhibitory is greater) avalanching dynamics can occur even for k/k_c substantially

above one.

4.3.2 Ceaseless Firing

For cases in which the connections between neurons are sufficiently dense and the fraction of excitatory neurons is sufficiently high, our model leads to ceaseless firings. Specifically, even in the absence of external drive, during ceaseless dynamics,

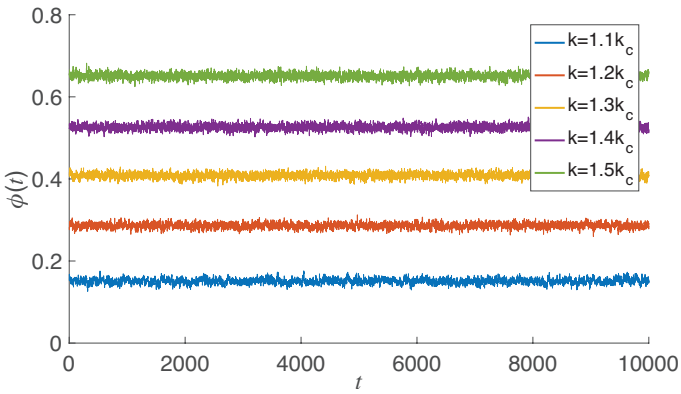


Figure 4.3: This figure shows the fraction of firing neurons versus time for simulations with ceaseless firing ($k > k_c$). Simulation is done on annealed random networks with $N = 10000$, threshold $M = 5$, and fraction of excitatory neurons $\alpha = 0.9$.

the fraction of firing neurons,

$$\phi(t) = N^{-1} \sum_{i=1}^{N} f_i(t), \tag{4.6}$$

is always positive. In Figs. 4.3, we plot the fraction of firing neurons versus time for 10^4 time steps, for cases in which the average degree of the network is greater than the predicted critical degree k_c. Note that Fig. 4.3 shows that, in this regime, $\phi(t)$ is approximately constant in time with only relatively small fluctuations, and avalanches do not occur.

4.3.3 Continuous Phase Translation

As the averaged degree k increases with other parameters (α, M) held fixed, the model behavior transitions from avalanche dynamics to ceaseless firing. Notice, that without the external drive, in the avalanche regime, the no-firing state $\phi(t) = 0$ is an absorbing state. We define an order parameter as the time averaged firing strength ϕ^* without the external drive (i.e., d_i set to 0 for all t in Eq. (4.3)),

$$\phi^* = \lim_{T \to +\infty} \frac{1}{T} \sum_{t=0}^{T} \phi(t). \tag{4.7}$$

As shown by Figs. 4.2 and 4.3, in the range of averaged degree k smaller than k_c, avalanches are observed, and thus, under the circumstance of no external drive, the order parameter is $\phi^* = 0$. According to our theory (Sec. 4.4), in the $N \to \infty$ limit, as k increases from k_c, ceaseless dynamics arise, and the order parameter ϕ^* becomes positive. We define the critical point at $k = k_c$ as the $N \to \infty$ limit of the border between avalanche dynamics and ceaseless firing dynamics. In Sec. 4.4 we employ a mean-field theory to obtain the stationary firing strength, as well as the condition on system parameters (k, α, and M) for criticality. (The values of k that we denote k_c, are from the analytical formula for k_c we derive in Sec. 4.4.)

4.3.4 Restored Avalanche for $k > k_c$

According to our definition of k_c, avalanche dynamics do not occur for $k > k_c$ in the $N \to \infty$ limit. We denote, however, that, for finite N, avalanches can occur.

For example, from our simulations with $N = 10^4$ we find that when k is just slightly above k_c, the system shows non-ceaseless firing avalanches in which the histogram of avalanche size is power-law below a certain size S but then ends with a wide tail. This can be understood by referring to Fig. 4.3 which shows that as k approaches k_c from above, the predicted ϕ^* approaches 0. Furthermore, Fig. 4.3 also shows fluctuations of $\phi(t)$ about ϕ^*. Thus, when ϕ^* is sufficiently small, the system has a positive chance of reaching the quiescent state, $\phi(t) = 0$, and the firing ceases instead of continuing as would be expected in the $N \to \infty$ limit. As a consequence, the simulations can occasionally exhibit much larger avalanches than those in critical and sub-critical cases ($k \leq k_c$), and thus wide tails can be observed in the avalanche size/duration distributions (e.g., Fig. 4.4).

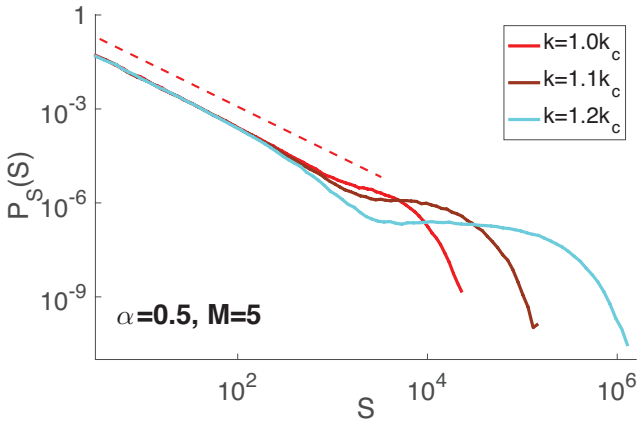


Figure 4.4: The firing avalanche size distribution with $M = 5$, $\alpha = 0.5$ for $k/k_c = 1.0$, 1.1, and 1.2. The two super-critical cases, $k/k_c = 1.1$ and 1.2, exhibit "restored" firing avalanches because the low mean-field predicted ϕ^* can not sustain ceaseless dynamics under fluctuations in $\phi(t)$. For $k/k_c \geq 1.3$ the system exhibits ceaseless dynamics.

We emphasis that the size of the range of k/k_c (in the super-critical region $k > k_c$) where "restored" avalanching appears depends on how fast ϕ^* increases from

0 as k/k_c is increased from 1, which is determined by α and M. For example, when $M = 5$ and $\alpha = 0.5$, the ϕ^* stays close to 0 for a large range of $k/k_c > 1$ (yellow curve in the third panel in Figs. 4.5) and thus restored avalanches are observed at $k = 1.1k_c$ and $k = 1.2k_c$ (Fig. 4.4).

In the next section we employ a mean-field analysis to study criticality of our model and explain the observed power law scaling in avalanche size and durations. In Sec. 4.5 we test our predictions for critical condition and the stationary firing strength ϕ^* in $k > k_c$. We also discuss the observed restoration of non-ceaseless firing avalanches with wide-tail size distributions when k is slightly above k_c. The robustness of critical power law behavior to deviations of k from $k = k_c$ is studied in Sec. 4.6.

4.4 Mean-Field Analysis of Criticality

To analytically explore criticality in our model, we develop a large N mean-field approach. With Erdős-Rényi type network topology, we assume that neurons from the full system can be treated as statistically identical, except that some are excitatory and some are inhibitory. We imagine that we probe the dynamics of the full system by studying a single hypothetical neuron that interacts with the mean-field. The single mean-field neuron's probability of being in state m at time t, denoted $P_m(t)$, corresponds to the fraction of neurons in state m at time t in the full system. Similarly, its firing probability at time t corresponds to the fraction of firing neurons at time t in the full system, and thus is denoted $\phi(t)$. At each time t,

this single mean-field neuron receives net excitatory and inhibitory signals from the mean-field, denoted by $s_+(t)$ and $s_-(t)$. Furthermore, we let $\theta(s)$ denote the time-dependent probability distribution function of the total signal $s(t) = s_+(t) - s_-(t)$, and $\underline{\theta}(x)$ and $\bar{\theta}(x)$ correspond to the probabilities that $s(t) \geq x$ and $s(t) \leq x$. Thus our mean-field description evolves as follows,

$$\begin{aligned}
\phi(t+1) &= \sum_{m=0}^{M} \underline{\theta}(M-m+1) P_m(t), && (4.8)\\
P_{m'}(t+1) &= \sum_{m=0}^{M} \theta(m'-m) P_m(t), \quad (m' \neq 0) && (4.9)\\
P_0(t+1) &= \sum_{m=0}^{M} \left[\bar{\theta}(-m) + \underline{\theta}(M-m+1)\right] P_m(t). && (4.10)
\end{aligned}$$

Considering $s_+(t)$ and $s_-(t)$ as independent random variables, we now separately derive their distribution functions and then obtain the distribution function $\theta(s)$ for the total signal s. At time t, the total number of firing neurons in the mean-field is $N\phi(t)$, of which $n_+ = \alpha N\phi(t)$ are excitatory and $n_- = (1-\alpha)N\phi(t)$ are inhibitory. We assume that the chance that our hypothetical mean-field neuron receives a signal from any firing neuron is independent of the firing of other neurons and has probability $p = k/N$. Thus, we consider the number of excitatory and inhibitory signals the mean-field neuron receives follow independent binomial distributions, $B(n_+, p)$ and $B(n_-, p)$. In the limit $N \to \infty$, with positive $\phi(t) > 0$ and $0 < \alpha < 1$, we expect $n_+ \gg 1$, $n_- \gg 1$, and $k \ll N$. Thus we further approximate

the binomial distributions by Poisson distributions with $\lambda_+ = n_+ p$ and $\lambda_- = n_- p$,

$$\Pr[s_+ = q; \lambda_+] = \frac{(\lambda_+)^q e^{-\lambda_+}}{q!}, \tag{4.11}$$

$$\Pr[s_- = q; \lambda_-] = \frac{(\lambda_-)^q e^{-\lambda_-}}{q!}. \tag{4.12}$$

Since the integers s_+ and s_- in our model are independent random variables, the probability distribution $\theta(s)$ of $s = s_+ - s_-$ is $\sum_{q_1} \sum_{q_2} Pr[q_1; \lambda_+] Pr[q_2; \lambda_-] \delta(s - q_1 + q_2)$ which is given in [84],

$$\theta(s) = e^{-k\phi(t)} \left(\frac{\alpha}{1-\alpha} \right)^{s/2} I_{|s|}(2\sqrt{\alpha(1-\alpha)} k\phi(t)), \tag{4.13}$$

where $I_{|s|}(\cdot)$ is the modified Bessel function of the first kind.

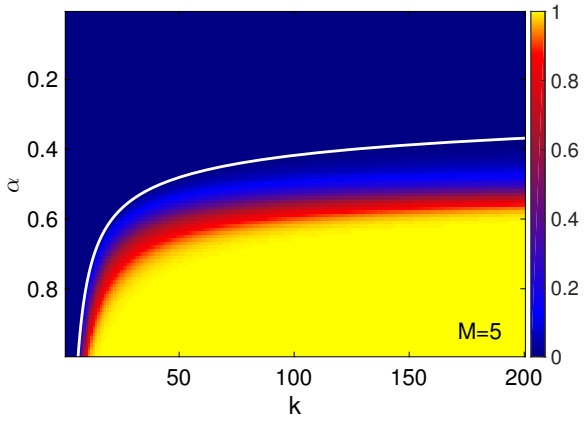


Figure 4.5: This figure shows the stationary firing strength $\phi^*(k, \alpha)$ predicted by the mean-field analysis (color) for different fraction of excitatory neurons α (vertical axis), average degree k (horizontal axis), with threshold value $M = 5$. The color bar shows the fraction of firing neurons when an equilibrium is reached, where dark blue indicates that there is no firing and light yellow indicates that all neurons are firing. The white curve indicate the critical dynamics which is the boundary between ceaseless firing and firing avalanches, predicted by Eqs. (4.23) and (4.25).

With the obtained distribution function $\theta(s)$, we numerically solve Eqs. (4.8)-

(4.10) for stationary solutions P_m^* and ϕ^*. Plots of stationary firing probability ϕ^* predicted by our mean-field method are shown in Figs. 4.5 with horizontal and vertical axises k and α for $M = 5$. A positive stationary ϕ^*, in light color, indicates that the firing of the network is ceaseless, while $\phi^* = 0$, colored in dark blue, indicates that the system cannot maintain firing without external drive, and thus avalanches can be seen when external drive during quiescence is included. When the averaged degree k and the fraction of excitatory neuron α are higher, ϕ^* is larger, in agreement with Fig. 4.3.

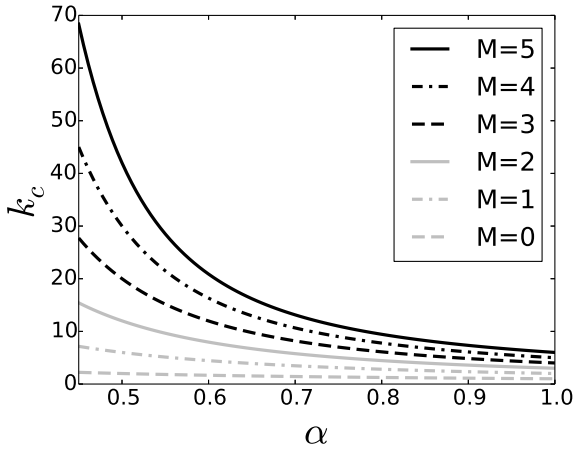


Figure 4.6: The curves show the critical average degree k_c versus the fraction of excitatory neurons α for $M = 0$ to $M = 5$, as predicted by Eqs. (4.23) and (4.25).

To obtain an analytic formula for the critical condition, we use the fact that, as we asymptotically approach the critical point from the super-critical region, ϕ^* becomes small. Expanding $\theta(s)$ from Eq. (4.13) around ϕ^*, to first order in ϕ^* we

get

$$\theta(s > 1; \alpha, k, \phi^*) = 0, \tag{4.14}$$

$$\theta(s = 1; \alpha, k, \phi^*) = \alpha k \phi^*, \tag{4.15}$$

$$\theta(s = 0; \alpha, k, \phi^*) = 1 - k\phi^*, \tag{4.16}$$

$$\theta(s = -1; \alpha, k, \phi^*) = (1 - \alpha) k \phi^*, \tag{4.17}$$

$$\theta(s < -1; \alpha, k, \phi^*) = 0, \tag{4.18}$$

which when inserted into to Eqs. (4.8)-(4.10) yields

$$\phi^* = \phi^* P_M^* \alpha k, \tag{4.19}$$

$$P_M^* = P_{M-1}^* \phi^* \alpha k + P_M^* \left(1 - k\phi^*\right), \tag{4.20}$$

$$P_m^* = P_{m-1}^* \phi^* \alpha k + P_{m+1}^* \phi^* (1 - \alpha) k + P_m^* \left(1 - k\phi^*\right), \tag{4.21}$$

$$P_0^* = P_M(t) \phi^* \alpha k + P_1^* \phi^* (1 - \alpha) k + P_0^* \left(1 - \phi^* \alpha k\right), \tag{4.22}$$

where $m = 1, 2, ..., M - 1$. Equations (4.19)-(4.20) suggest that $P_M^* = 1/(k\alpha)$ and $P_{M-1}^* = 1/(k\alpha^2)$, while the P_m^* for $0 \leq m < M - 1$ can be solved for in terms of α, k, and m from Eqs. (4.21) and (4.22). Then, by using the normalization property

$\sum_{m=0}^{M} P_m = 1$, we obtain the critical degree k_c as well as the P_m^*,

$$k_{\rm c} = \frac{(2M+3)\alpha^2 - (M+2)\alpha + (\alpha-1)^2(\alpha^{-1}-1)^M}{\alpha(1-2\alpha)^2}, \tag{4.23}$$

$$P_m^* = \frac{(2\alpha-1)\left[\alpha + (\alpha-1)(\alpha^{-1}-1)^{M-m}\right]}{(2M+3)\alpha^2 - (M+2)\alpha + (\alpha-1)^2(\alpha^{-1}-1)^M}, \tag{4.24}$$

when $\alpha \neq 0.5$, and

$$k_{\rm c} = (M+1)(M+2), \tag{4.25}$$

$$P_m^* = \frac{2(M-m+1)}{(M+1)(M+2)}. \tag{4.26}$$

when $\alpha = 0.5$. The critical condition is plotted as white curves in Figs. 4.5 that separate the avalanche region and the ceaseless firing region. The critical average degree k_c as a function of α for different M is plotted in Fig. 4.6. In agreement with intuition, the critical average degree becomes higher as more inhibitory neurons are included.

4.5 Simulation Results and Comparison with Mean-Field Analysis

We performed simulation of random networks with $N = 10^4$ neurons and random initial conditions $\{x_i(t)\}$. To obtain histograms for avalanche sizes and durations ($k \leq k_c$), during each simulation, we collected 10^6 avalanches. For firing activity lasting more than $2N$ time steps, we consider the dynamics to be ceaseless.

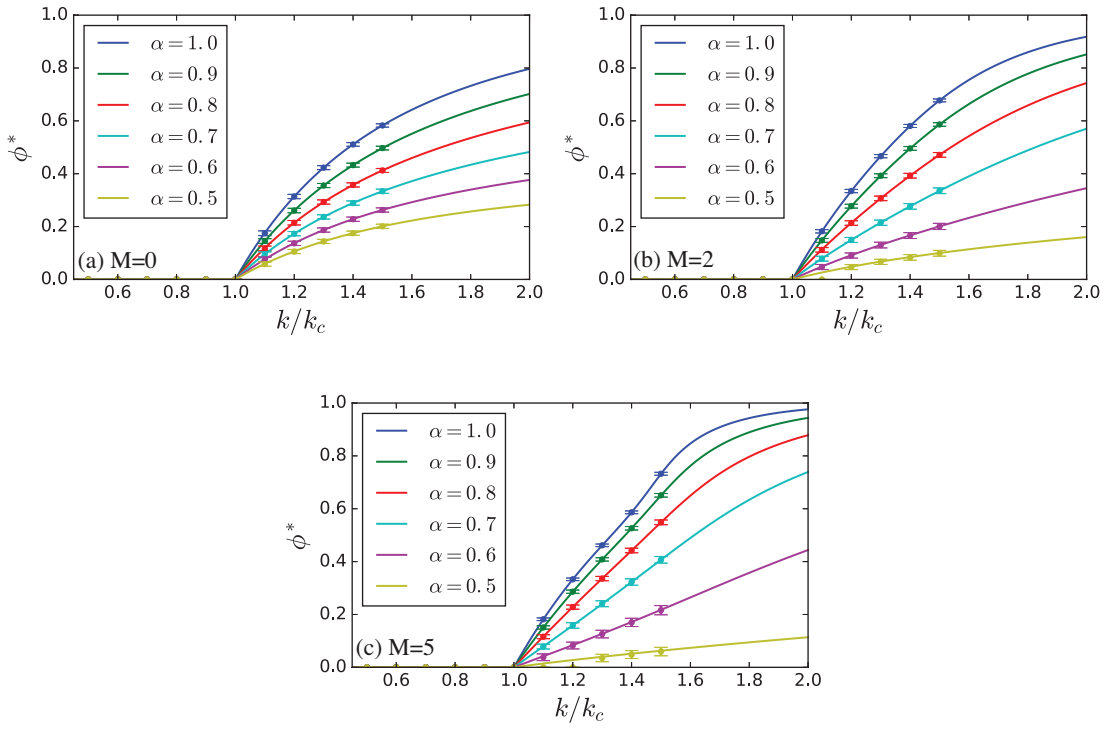


Figure 4.7: Comparison of the mean-field theoretically predicted ϕ^* (curves) with simulation results (points with error bars) for annealed random networks for $M = 0, 2, 5$. The error bars indicate the fluctuations (standard deviations) of the fraction of firing neurons $\phi(t)$ during ceaseless firing dynamics.

In Figs. 4.7 and 4.8 we compare the time averaged stationary firing strengths ϕ^* predicted by theory with those observed in simulations.

In Fig. 4.7 we plot the mean-field predicted stationary firing strength, ϕ^*, as a function of k/k_c for different M and α as well as the time averaged firing strength observed from simulation (dots with error bars indicating standard deviations), for annealed random Erdős-Rényi network with $M = 0, 2, 5$ and $\alpha = 1, 0.9, 0.8, 0.7, 0.6, 0.5$. We see that for $k/k_c > 1$, the average firing strength is lower for lower α, and this effect is enhanced for larger M.

In order to better illustrate how the simulation results differ from the mean-field predictions, in Fig. 4.8 we show the difference between the firing strength as determined from the simulations, ϕ^*_{sim} (squares: annealed, circles: partially annealed, triangles: quenched) and the value determined from the mean-field theory, ϕ^*_{th}. In the annealed case, there is very good agreement between the simulations and the theory. This is expected as the theory's assumption of independence between different neuron firings is valid for the annealed case. We observe that there is some disagreement with the mean-field theory for the partially annealed and quenched cases, with (as expected) the partially-annealed case showing better agreement than the quenched case.

Plotting ϕ^* versus k/k_c for $M = 5$ (as in the third panel of Fig. 4.7), we find that ϕ^* for $k/k_c > 1$ can be quite small. Noting the fluctuations of $\phi(t)$ about its mean (ϕ^*) seen in Fig. 4.3, the mean-field predicted size of ϕ^* for $k/k_c > 1$ at $M = 5$, $\alpha < 0.5$ can be less than the fluctuation width. Thus, for finite N, $\phi(t)$ can fluctuate down to zero, and the dynamics are not ceaseless as predicted by the

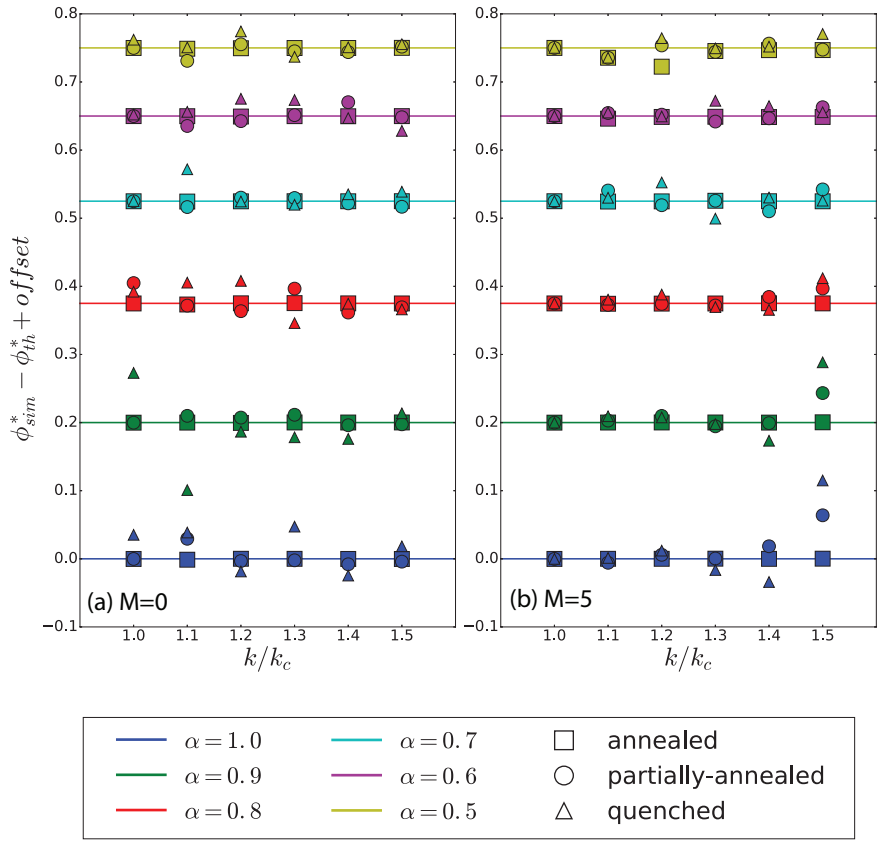


Figure 4.8: Deviations between the stationary ceaseless firing strength predicted by the mean-field analysis, ϕ^*_{th}, and those obtained from the simulation results, ϕ^*_{sim}, for annealed (squares), partially annealed (circles), and quenched (triangles) random networks are plotted for $M = 0$ in (a) and $M = 5$ in (b) with different offsets for different α values. Different colors indicate different values of α. The distance between markers and colored horizontal lines indicate the deviations between simulation and theory. (For the purpose of visualization, the offsets are not evenly spaced.)

mean-field theory. In such a situation, avalanches will be restored (Fig. 4.4). As we see in the next section, this is a very significant feature that helps support robust criticality.

4.6 Robustness of Criticality

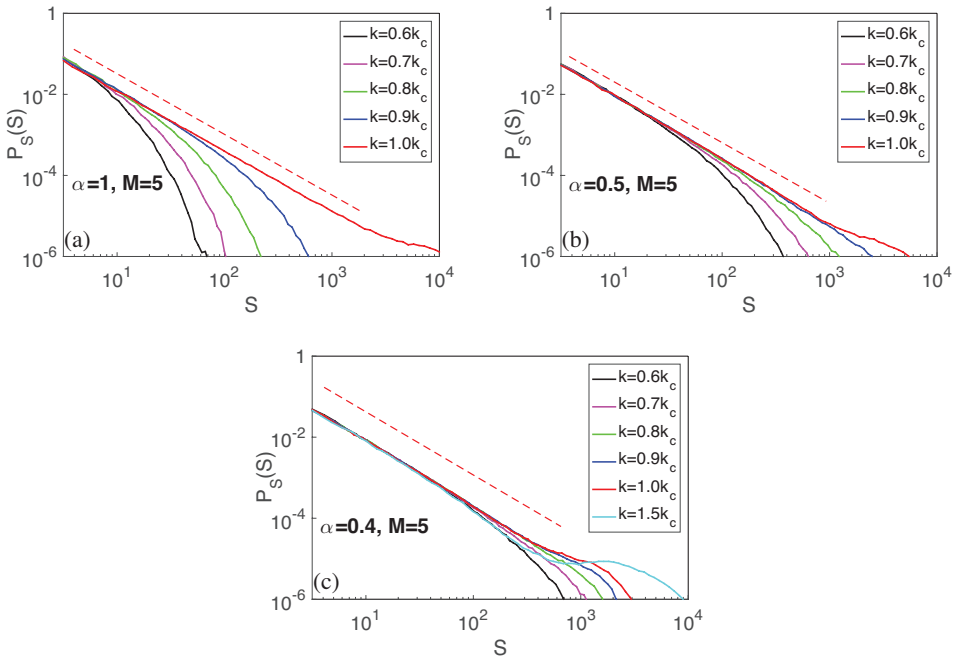


Figure 4.9: Histogram of avalanche size distributions with $M = 5$, $\alpha = 1,\ 0.5,\ 0.4$, and several different degrees. The red dashed line indicate the power law exponent $-3/2$. As α is lowered, the size distributions become more and more close to the power law, indicating an increment of robustness of criticality.

In our study, we observe enhanced robustness of criticality as more inhibitory neurons are included (i.e., as α is decreased), and this enhancement is more significant for cases with larger M (i.e., the discrete dynamics have stronger integrate-and-fire type behavior).

In Figs. 4.9(a)-(c), we plot avalanche size (S) histograms, for k from $0.6k_c$ to

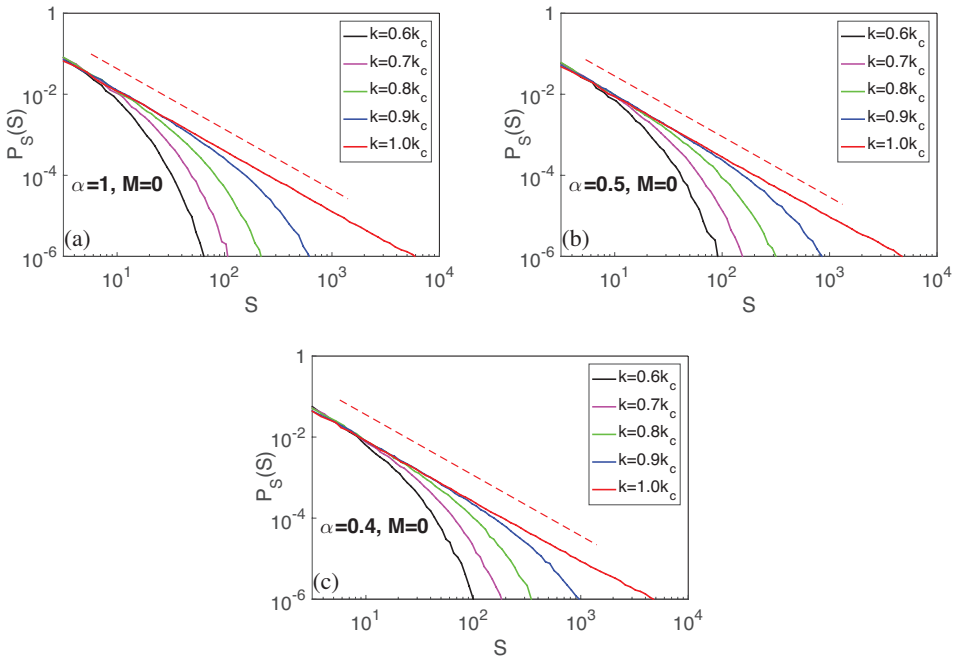


Figure 4.10: Histogram of avalanche size distributions with $M = 0$, $\alpha = 1,\ 0.5,\ 0.4$, and several different degrees. The red dashed line indicate the power law exponent $-3/2$. The three panels looks similar and the increment of robustness is not as obvious as that shown in Fig. 4.9.

$1.0k_c$, $\alpha = 1$, 0.5, and 0.4, with $M = 5$. As we mentioned above, when $k = k_c$, the avalanche size distribution exhibits the power law with exponent $-3/2$. When $k < k_c$, the histogram bends downward, but may stay close to the critical ($k = k_c$) behavior for low S, and this range of proximity increase as k/k_c approaches one. Examining Figs. 4.9 (a)-(c), we see that smaller values of α promote enhanced robustness of critical behavior. We also notice that when $\alpha = 0.4$, due to the fact that the predicted ϕ^* is very small for a significantly larger range of $k/k_c > 1$, restored avalanches can be observed for large k. In Fig. 4.9(c) we plot the avalanche distribution at $k = 1.5k_c$ in cyan which approximately follows the power law for about three decades in S and ends with a fat tail. It is remarkable that the clear $-3/2$ power law behavior persists for $\alpha = 0.4$, $M = 5$ over a range in k by a factor

of at least $1.5/0.6 = 2.5$. (When $\alpha = 0.4$, and $M = 5$, restored avalanches can still be seen for $k > 1.5k_c$; we only plot up to $k = 1.5k_c$, however, due to the increased computing resources required for simulating very large avalanches.)

Examining Figs. 4.10 (a)-(c), we see that when $M = 0$, smaller values of α do not obviously promote enhanced robustness of critical behavior, compared with $M = 5$. Thus the enhancement of robustness due to the inhibitory neurons is only significant when the dynamics is more of the integrate-and-fire type.

To confirm and visualize the effect of inhibitory-neuron-enhanced robustness of criticality under perturbation of the average degree k away from k_c, we determine the span in $\log_{10}(S)$ over which the slope of the avalanche size distribution in the log-log plot (as approximated from the numerical data) stays close enough (± 0.3) to the critical exponent $-3/2$. We note this span by $\Delta(k)$ (see Fig. 4.11(a) and its caption for further details). For a given value of $k \neq k_c$, the ratio of this value to the one observed for $k = k_c$ provides a measure of similarity to criticality for the specified parameter values:

$$\rho(k, \alpha, M) = \frac{\Delta(k; \alpha, M)}{\Delta(k_c; \alpha, M)}. \tag{4.27}$$

In Fig. 4.11(b), we plot $\rho(0.9k_c, \alpha, M)$ versus α and M. From this plot, we see that the inclusion of inhibitory neurons (smaller α) enhances the robustness of criticality (i.e., ρ near 1 for non-critical parameters), and, when M is larger (i.e., stronger integrate-and-fire type dynamics), this robustness-enhancing effect is stronger than

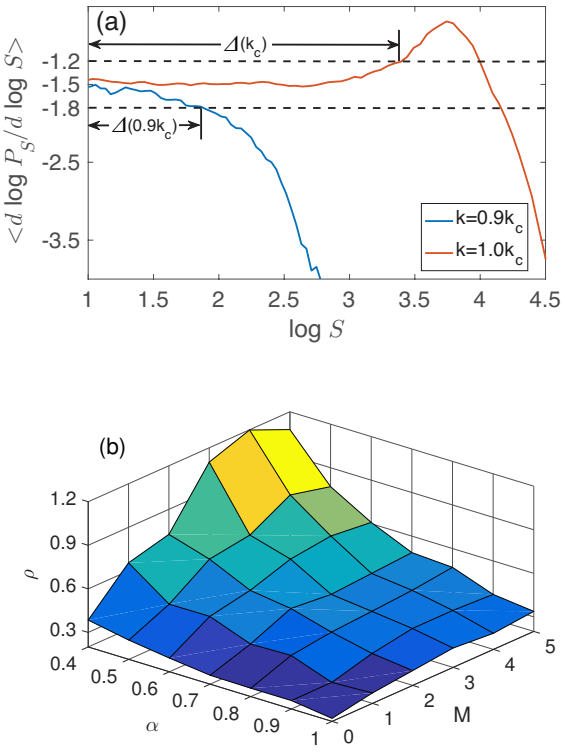


Figure 4.11: (a) This figure illustrates how we determine $\Delta(k)$ using the example cases $k = k_c$ and $k = 0.9k_c$ at $\alpha = 0.9$, $M = 5$. We first determine an approximation to $d\log_{10} P_S/d\log S$ by evaluating P_S from data like that in Figs. 4.9 and 4.10 at $\log S + \delta$ and $\log S - \delta$. We choose δ to be small compared to one, yet large enough to acceptably reduce statistical fluctuations in the difference of the values of P_S at $\log S + \delta$ and $\log S - \delta$. Specifically, we set $\delta/[\log S_{max} - \log S_{min}] = 0.04$. We then take as our slope estimate, denoted $< d\log P_S/d\log S >$, this difference in P_S divided by 2δ. We then define $\Delta(k)$ to be the range in $\log S$ (starting at $\log S = 1$) over which the estimated slope is within a distance 0.3 of the expected critical slope -1.5. (b) $\rho(0.9k_c; \alpha, M) = \Delta(0.9k_c; \alpha, M)/\Delta(k_c; \alpha, M)$ versus α and M.

for lower M [2].

4.7 Conclusions

In this chapter, we study the firing dynamics of a discrete-state/discrete-time model of integrate-and-fire type neural network dynamics. Both firing avalanches and ceaseless firing are observed with a continuous phase transition connecting these two regimes. By using a mean-field method, we find an analytical condition for criticality and predict the average firing strength in the ceaseless firing regime. We find that the theoretical predictions are in good agreement with simulation results from annealed, partially-annealed, and quenched Erdős-Rényi random networks, with best agreement with annealed networks. Most importantly, we have also found that the robustness of critical behavior is enhanced by the simultaneous presence of a large fraction of inhibitory neurons (i.e., α near or less than 0.5) and integrate-and-fire type dynamics (i.e., sufficiently large M).

[2]For $\alpha < 0.5$, we have $\rho \leq 1$ because the subcritical size distribution deviates from critical at lower S, as k/k_c is decreased from one. ρ values larger than one at the corner where $M = 5$ and $\alpha = 0.4$ can be explained as follows. As shown in Fig. 4.9(c), the critical avalanche size distribution deviates from the power law by bending upward and then downward forming a large S tail to P_S. Due to the simultaneous canceling tendencies for increasing slope (upward and bend) and decreasing slope (downward bend) for the $k = 0.9k_c$ curve (see Fig. 4.9 third panel for $\log_{10} S$ near 3.3) the range of $\log_{10}(S)$ that seems to follow the $-3/2$ power law is longer than the corresponding range for $k = k_c$.

Chapter 5: Reservoir Observers: Model-free Inference of Unmeasured Variables in Chaotic Systems

This chapter is based on work contained in the publication: Reservoir observers: Model-free inference of unmeasured variables in chaotic systems, Lu, Z., Pathak, J., Hunt, B., Girvan, M., Brockett, R., and Ott, E., Chaos, 27(4), 041102 (2017), © 2017 by the American Institute of Physics.

5.1 Introduction

Frequently, when studying the dynamics of a physical system, one only has access to a limited set of measurements of the state variables and desires to deduce concurrent values of unmeasured state variables. In principle, it might be possible to accomplish this goal if, in addition to the measurements, one also has knowledge of the system dynamics. In control theory, a successful deduction method of this type is called an *observer*. Observers are of great utility for control and prediction of dynamics. The observer problem for the case in which the dynamical system is linear was fully solved in the classic work of Kalman [85], who also formulated conditions for "observability" under which it is possible to achieve the goal of deducing the full state of a linear system from a given partial set of state measurements (see textbooks

on control theory, e.g., Ref. [86]). Observers and observability have also been extensively investigated for nonlinear dynamical systems (e.g., Ref. [87]). For example, in situations of chaotic dynamics, an approach using synchronization of chaos has been exploited [88, 89]. A concept related to observability, called "estimability" has been discussed in the context of delay coordinate embedding [90].

In this chapter we consider the observer problem for situations in which one does not have a sufficiently accurate mathematical model of the nonlinear system of interest. In place of such a model, we assume that there exists an initial period of time for which measurements of all the desired system variables are available, and we seek to use these measurements in the initial period of time to deduce the full set of desired variables for the subsequent time, for which we assume that measurements of only a limited subset of the desired variables are possible. Our method utilizes a machine learning technique, called *reservoir computing* (see Ref. [28]). This technique employs an input/output neural network with randomly generated parameters, and uses linear regression to choose "output weights" that fit the raw network output to a set of "training data". We use the data from the initial period of full measurement as the training data. Then we continue to input the subsequent partial set of continually measured variables, and regard the weighted network output as the estimated current values of the variables that are no longer measured. (We emphasize that the goal we address is the inference of unmeasured state variables rather than their prediction.)

The main result of this chapter is that this kind of "reservoir observer," subject to certain limitations, can be extremely effective. In what follows we first de-

scribe a specific illustrative implementation and review relevant reservoir computing concepts. We then discuss applications to three examples of chaotic systems that highlight the strength and limitations of our method: (1) the Rössler system [91], for which we have done an intensive study of how results depend on design parameters of the reservoir observer; (2) the Lorenz system [92], which we use to illustrate an instance of the issue of observability for our method.

5.2 Setup

We consider a dynamical system $d\boldsymbol{\phi}/dt = \mathbf{f}(\boldsymbol{\phi})$ together with a pair of $\boldsymbol{\phi}$-dependent, vector valued variables, $\mathbf{u} = \mathbf{h_1}(\boldsymbol{\phi}) \in \mathbb{R}^M$ and $\boldsymbol{s} = \mathbf{h_2}(\boldsymbol{\phi}) \in \mathbb{R}^P$. We are interested in the situation in which $\mathbf{u}$ and $\mathbf{s}$ can both be measured over a specific period, $[0, T]$, but that only $\mathbf{u}$ can be measured from that time forward; we seek a method for using the continued knowledge of $\mathbf{u}$ to determine an estimate of $\mathbf{s}$ as a function of time when direct measurement of $\mathbf{s}$ is not available, $t > T$. In contrast with most of the engineering literature devoted to problems of this kind, we do not assume knowledge of $\mathbf{f}$ but rather seek to infer the necessary information from the trajectories recorded on the interval $[0, T]$.

For this purpose we use "reservoir computing" [28], which has previously been advocated for application to many tasks (e.g., prediction of time series, pattern recognition, etc.). There are many variations in implementation; in this chapter we adopt the reservoir technique proposed by Jaeger [93]. The reservoir computer has three components (Fig. 5.1), a linear input layer with M input nodes (one for

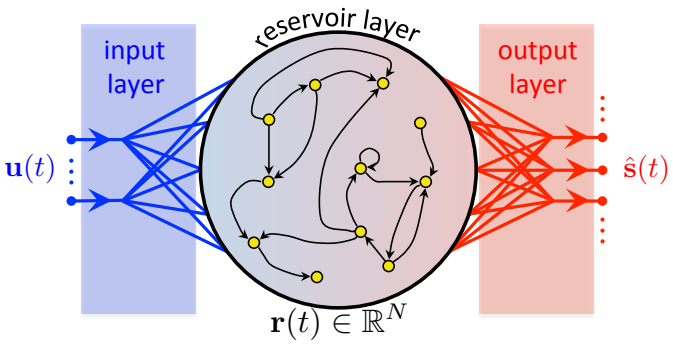


Figure 5.1: A reservoir computer consisting of three parts, an input layer, a reservoir layer with state $\mathbf{r}(t)$, and an output layer. For $t > T$, the input to the system is $\mathbf{u}(t)$ and our goal is that the output $\hat{\mathbf{s}}(t)$ is a good approximation to the unmeasured quantity $\mathbf{s}(t)$.

each component of $\mathbf{u}$), a recurrent, nonlinear reservoir network with N dynamical reservoir nodes whose state vector is $\mathbf{r} \in \mathbb{R}^N$, and a linear output layer with P output nodes, as shown in Fig. 5.1. For the specific reservoir computing implementation we use in this chapter, the reservoir dynamics is defined as

$$\mathbf{r}(t + \Delta t) = (1 - \alpha)\mathbf{r}(t) + \alpha \tanh(\mathbf{A}\mathbf{r}(t) + \mathbf{W}_{in}\mathbf{u}(t) + \xi\mathbf{1}), \tag{5.1}$$

where we assume the time step to be small $\Delta t \ll 1$, $\mathbf{A}$ is the (typically sparse) weighted adjacency matrix of the reservoir layer, and the M-dimensional input $\mathbf{u}(t)$ is fed in to the N reservoir nodes via a linear input weight matrix denoted by $\mathbf{W}_{in} \in \mathbb{R}^{N \times M}$. (In Secs. 5.3.1 and 5.3.2, $M = 1$ since the input $\mathbf{u}$ is a scalar.) The parameter $0 < \alpha \leq 1$ is a "leakage" rate [94] that makes the reservoir evolve more slowly as $\alpha \to 0$. We also use a bias term $\xi\mathbf{1}$, where $\mathbf{1}$ denotes a vector of ones and ξ is a scalar constant. The notation $\tanh(\cdot)$ with a vector argument is defined as the vector whose components are the hyperbolic tangents of the corresponding

components of the argument vector. The output, which is a P-dimensional vector, is taken to be a linear function of the reservoir state,

$$\hat{\mathbf{s}}(t) = \mathbf{W}_{out}\mathbf{r}(t) + \mathbf{c}. \tag{5.2}$$

As compared to other artificial neural network approaches, the advantage of reservoir computing is that training is made computationally feasible for relatively large N, since only the output weights $\mathbf{W}_{out}$ and the vector $\mathbf{c}$ are adjusted by the training process. (The input weight matrix $\mathbf{W}_{in}$ and the reservoir adjacency matrix $\mathbf{A}$ are initially randomly drawn and then fixed.) A key point is that the reservoir layer serves as an active medium driven by inputs $\mathbf{u}(t)$ where each reservoir node has a different nonlinear response to its inputs, so that for $N \gg 1$ we can hope that almost a wide variety of desired outputs can be approximated by a linear combination of the N-dimensional reservoir nodal response states.

In addition to the parameters Δt, α, and ξ in Eq. (5.1), and the reservoir size N, the reservoir dynamics depend on the parameters D, ρ, and σ, which govern the random generation of $\mathbf{A}$ and $\mathbf{W}_{in}$ as follows. The adjacency matrix $\mathbf{A}$ is built from a sparse random Erdős-Rényi matrix in which the fraction of nonzero matrix elements is D/N, so that the average degree of a reservoir node is D. The values of non-zero elements are randomly drawn independently from a uniform distribution between -1 and 1. We then uniformly rescale all the elements of $\mathbf{A}$ (i.e., multiply $\mathbf{A}$ by a positive scalar) so that the largest value of the magnitudes of its eigenvalues becomes ρ, which we refer to as the "spectral radius" of $\mathbf{A}$. For the input layer, the

ith of the M input signals is connected to N/M reservoir nodes with connection weights in the ith column of $\mathbf{W}_{in}$. Each reservoir node receives input from exactly one input signal. The non-zero elements of $\mathbf{W}_{in}$ are randomly chosen from a uniform distribution in $[-\sigma, \sigma]$. (We emphasize that this choice of structure for $\mathbf{W}_{in}$ is fairly arbitrary and that other choices could reasonably be employed, e.g., every node could receive a mixture of several inputs.)

For the convenience of comparing the reservoir performances, we preprocess all the components of $\mathbf{u}(t)$ and $\mathbf{s}(t)$ so that they have zero mean and unit variance. Starting from a random initial state $\mathbf{r}(-\tau)$, the reservoir evolves following Eq. (5.1) with input $\mathbf{u}(t)$. Here τ is a transient time, chosen large enough to make the reservoir state essentially independent of its initial state by time $t = 0$. We then record the $K = T/\Delta t$ reservoir states for $0 < t \leq T$,

$$\{\mathbf{r}(\Delta t), \mathbf{r}(2\Delta t), ..., \mathbf{r}(T)\}, \tag{5.3}$$

and the concurrent measurements of the state variables that are unmeasured for $t > T$,

$$\{\mathbf{s}(\Delta t), \mathbf{s}(2\Delta t), ..., \mathbf{s}(T)\}. \tag{5.4}$$

We then train the network by choosing the output layer quantities $\mathbf{W}_{out}$ and $\mathbf{c}$ by choosing them so that the reservoir output approximates the measurement for $0 < t \leq T$. We do this by minimizing the following quadratic form with respect to

$\mathbf{W}_{out}$ and $\mathbf{c}$,

$$\{\sum_{k=1}^{K} \|\mathbf{W}_{out}\mathbf{r}(k\Delta t) + \mathbf{c} - \mathbf{s}(k\Delta t)\|^2\} + \beta[\mathrm{Tr}(\mathbf{W}_{out}\mathbf{W}_{out}^T)], \tag{5.5}$$

where $\|\mathbf{q}\|^2 = \mathbf{q}^T\mathbf{q}$ for $\mathbf{q}$ a vector. The second term of Eq. (5.5), $\beta[\mathrm{Tr}(\mathbf{W}_{out}\mathbf{W}_{out}^T)]$, is a regularization term included to avoid overfitting $\mathbf{W}_{out}$, where $\beta > 0$ (typically a small number) is the "ridge regression parameter".

If the training is successful, the readout of the reservoir output should yield a good approximation (denoted $\hat{\mathbf{s}}(t)$) to the desired unmeasured quantity $\mathbf{s}(t)$ for $t > T$. Referring to Eq. (5.2),

$$\hat{\mathbf{s}}(t) = \mathbf{W}_{out}^*\mathbf{r}(t) + \mathbf{c}^*, \tag{5.6}$$

where $\mathbf{W}_{out}^*$ and $\mathbf{c}^*$ denote the solutions for the minimizers of Eq. (5.5),

$$\mathbf{W}_{out}^* = \delta\mathbf{S}\delta\mathbf{R}^T(\delta\mathbf{R}\delta\mathbf{R}^T + \beta\mathbf{I})^{-1}, \tag{5.7}$$

$$\mathbf{c}^* = -[\mathbf{W}_{out}^*\bar{\mathbf{r}} - \bar{\mathbf{s}}], \tag{5.8}$$

$$\bar{\mathbf{r}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{r}(k\Delta t),\ \bar{\mathbf{s}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{s}(k\Delta t) \tag{5.9}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix, $\delta\mathbf{R}$ (respectively, $\delta\mathbf{S}$) is the matrix whose kth column is $\mathbf{r}(k\Delta t) - \bar{\mathbf{r}}$ (respectively, $\mathbf{s}(k\Delta t) - \bar{\mathbf{r}}$).

We remark that Eq. (5.1) represents a special choice for the form of the reservoir that is convenient for our purposes. More generally Eq. (5.1) can be expressed as

$$\mathbf{r}(t + \Delta t) = \mathbf{g}(\mathbf{r}(t), \mathbf{u}(t)), \tag{5.10}$$

and other forms of $\mathbf{g}$, different from the choice Eq. (5.1), have been employed for reservoir methods designed to implement goals different from the observer goal that we address here. For example, experimental reservoir implementations have been reported where the dynamics Eq. (5.10) were from an optical network of semiconductor lasers [95], a delay system with a single nonlinear node [96], a field-programmable gate array (FPGA) [97], phase-delay electro-optic devices [98], and even a bucket of water [99], among others. Such choices might also work for a reservoir-based observer and may offer advantages such as the potential for huge increase in speed [98]. The main requirement on the observer dynamics Eq. (5.10) seems to be that it is sufficiently complex and that the dimension of the reservoir state vector $\mathbf{r}$ is sufficiently large that the output Eq. (5.2) can be made to approximate the desired time series (for our goal, $\mathbf{s}(t)$) by adjustment of $\mathbf{W}_{out}$ and $\mathbf{c}$.

5.3 Examples

In this section, we illustrate the reservoir-based observer method of Sec. 5.2 using two chaotic dynamical systems: the Rössler system and the Lorenz system.

5.3.1 Rössler System

We now test the observing method given in Sec. 5.2 on a chaotic Rössler system,

$$dx/dt = -y - z, \tag{5.11}$$

$$dy/dt = x + ay, \tag{5.12}$$

$$dz/dt = b + z(x - c), \tag{5.13}$$

where $a = 0.5$, $b = 2.0$, and $c = 4.0$. We choose one of the coordinates of the system as the "continually available" measurement $\mathbf{u}(t)$ defined in Sec. 5.2 (i.e., $M = 1$). The other two coordinates are taken to be the components of $\mathbf{s}(t)$, knowledge of which is only available for the training phase $0 \leq t \leq T$.

We find that a trained reservoir computing network with input from any one of these three coordinates, x, y, and z, can generate output that accurately agrees with the two remaining coordinates. As shown in Figs. 5.2, a reservoir computer whose preprocessed input is

$$\tilde{x}(t) = [x(t) - \langle x(t) \rangle]/\langle [x(t) - \langle x(t) \rangle]^2 \rangle^{1/2} \tag{5.14}$$

can accurately output the similarly defined versions $(\tilde{y}(t),\ \tilde{z}(t))$ of $(y(t), z(t))$ for $t > T$, i.e., $\mathbf{u}(t) = \tilde{x}(t)$, $\mathbf{s}(t) = [\tilde{y}(t), \tilde{z}(t)]^T$. (In Eq. (5.14) the angle brackets denote time average.)

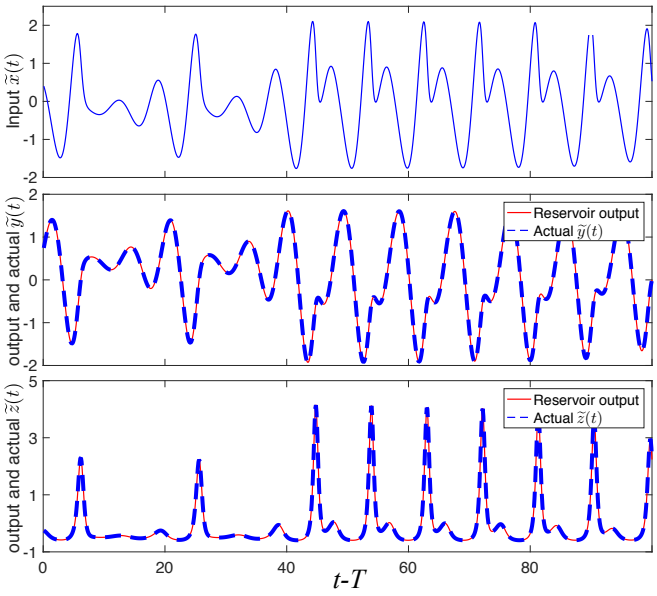


Figure 5.2: For the Rössler system Eqs. (5.11)-(5.13) with input $\tilde{x}(t)$, the reservoir computer very accurately generates $\tilde{y}(t)$ and $\tilde{z}(t)$. The reservoir parameters we use here are given in Eqs. (5.15).

For the results in Figs. 5.2 we utilize the following set of reservoir parameters (defined in Sec. 5.2),

$$\begin{aligned}
\text{number of reservoir nodes:} \quad & N = 400, \\
\text{spectral radius:} \quad & \rho = 1.0, \\
\text{average degree:} \quad & D = 20, \\
\text{scale of input weights:} \quad & \sigma = 1.0, \\
\text{bias constant:} \quad & \xi = 1.0, \\
\text{leakage rate:} \quad & \alpha = 1.0, \\
\text{time interval:} \quad & \Delta t = 0.1, \\
\text{length of training phase:} \quad & T = 260.
\end{aligned} \tag{5.15}$$

For this set of parameters the difference between the inferred (thin line in Figs. 5.2(b)

and 5.2(c)) and actual (dashed line for Figs. 5.2(b) and 5.2(c)) values of $\tilde{y}(t)$ and $\tilde{z}(t)$ is very small.

We examined how the root mean square (RMS) error of the inferred values varies with the reservoir parameters. Results of this type are given in Figs. 5.3(a)-5.3(f) where we plot the RMS error over 500 time units in the inference of $\tilde{y}(t)$ from the measurement of $\tilde{x}(t)$ versus one of the parameters ($\rho, D, \sigma, \xi, \alpha,$ or N), with the other parameters held fixed as given in Eqs. (5.15). Each plotted point in these figures is the median of 100 trials using the same signals $\tilde{x}(t)$ and $\tilde{y}(t)$, with each trial using an independent random realization of the reservoir system, and the error bars indicate the range in the performance of the observer from the second to the third quartile. Except for plots (b) and (f), the same 100 random choices are used for each value of the parameter. For example, Fig. 5.3(a) is a plot of the RMS error versus ρ. There is a wide range for which the error is small ($\lesssim 10^{-2}$), thus demonstrating that good observer performance is robust to significant changes in ρ. The plots in Figs. 5.3(a)-5.3(f) show that good performance is robust to changes in the other parameters, as well.

Motivated by the frequent consideration [100–102] of time delayed measurements for the analysis of experimental data from chaotic systems, we have also tested incorporation of time delayed versions of the input into the output by adding the term,

$$\sum_{k=0}^{K} d_k \mathbf{u}(t - k\Delta t), \tag{5.16}$$

into the reservoir output Eq. (5.2) and correspondingly also into the expression

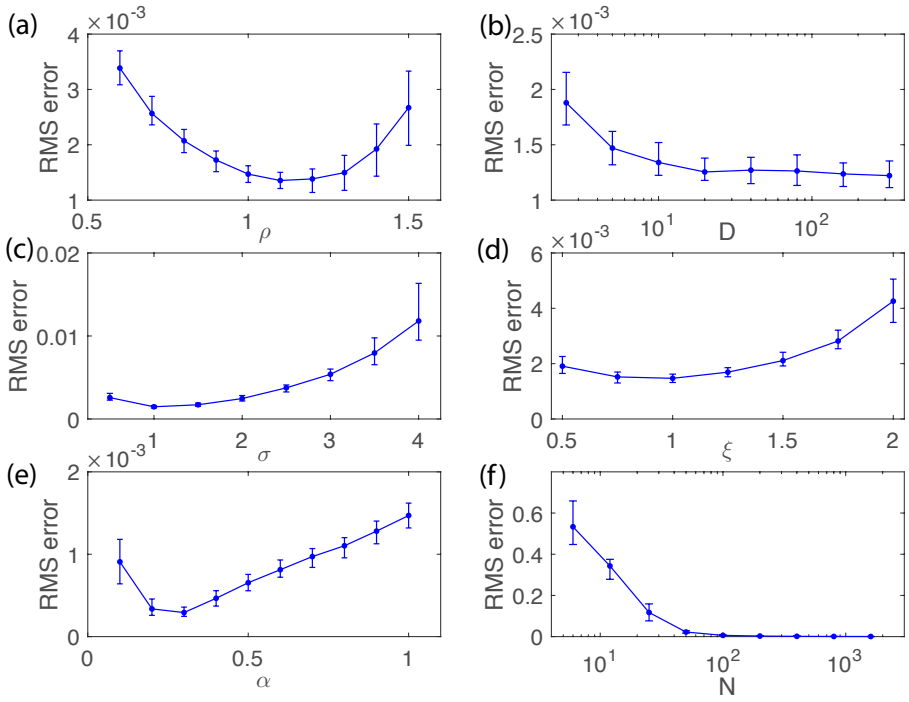


Figure 5.3: The RMS error in the inference of $\tilde{y}(t)$ from the measurement of $\tilde{x}(t)$ versus one of the parameters ($\rho, D, \sigma, \xi, \alpha,$ or N), with the other parameters held fixed as given in Eqs. (5.15). For the data in (f) with the RMS error on a logarithmic scale, see the curve in Fig. 5.4(a) labeled "without Eq. (5.16)."

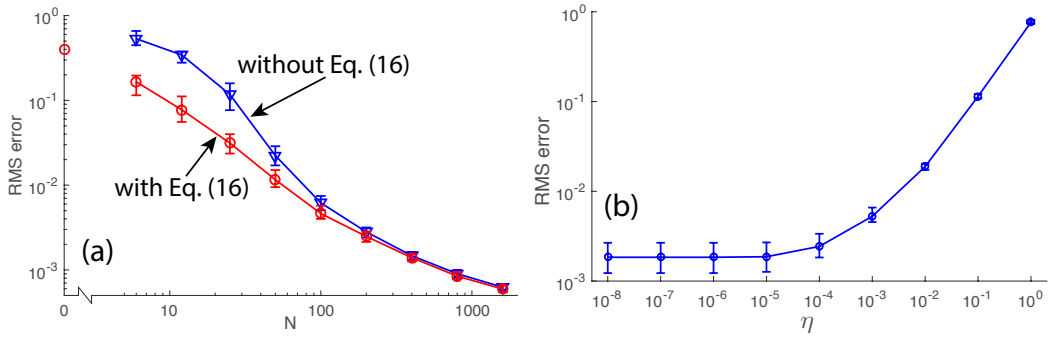


Figure 5.4: (a) The RMS error of the inference of $\tilde{y}(t)$ from $\tilde{x}(t)$ versus the size of the reservoir N with Eq. (5.16) and $K = 20$ and without Eq. (5.16). Aside from N, all the parameter values used in this figure are as given in Eqs. (5.15). (b) The effect of observational noise η in both $\tilde{x}(t)$ and $\tilde{y}(t)$ on the RMS error. The reservoir parameters for this plot are as given in Eq. (5.15). For both figures, the plotted points indicate the median error over 100 random realizations of the reservoir and the error bars indicate the range of the second and the third RMS error quartiles.

given by Eq. (5.5). We then minimize the expression in Eq. (5.5) over $\mathbf{W}_{out}$, $\mathbf{c}$, and $\mathbf{d} = (d_1, d_2, ..., d_K)^T$ to obtain our approximation $\hat{\mathbf{s}}$ for $\mathbf{s}$. An illustrative result is shown in Fig. 5.4(a), where we plot the median RMS error in $\tilde{y}$ inferred from $\tilde{x}(t)$ as a function of N for the two cases where Eq. (5.16) is included ($K = 20$ for Fig. 5.4(a)) and not included. We find that including delays does not significantly improve performance for $N > 100$, but does make some improvement when the number of reservoir nodes N is small and the error is relatively larger. For example, if one requires the error to be less than about 3×10^{-2}, then one can use as few as $N \sim 20$ reservoir nodes when the delays are included, while, without delays, $N \sim 50$ is required. We note that for $N = 0$ (no reservoir), the procedure reduces to a standard linear autoregressive approximation of $\mathbf{s}$. We also tested other values of K and found that the quantitative results were very much the same as for $K = 20$ (plotted as open circles in Fig. 5.4(a)) as long as $K > 5$.

Figure 5.4(b) shows the effect of observational noise in the input $\tilde{x}(t)$ and the training data $\tilde{y}(t)$, where, at every time step $\Delta t = 0.1$, observational noise was simulated by adding an independent random number with uniform distribution between $-\eta$ and $+\eta$ to each "measurement". We see that the noise has little influence on the error for $\eta \leq 10^{-4}$.

5.3.2 Lorenz System

In this section we test our method on the Lorenz system. This example provides insight into how the issue of 'observability' manifests itself for our method.

The Lorenz system is described by the equations,

$$\begin{aligned} dx/dt &= -ax + ay, \\ dy/dt &= bx - y - xz, \\ dz/dt &= cz + xy, \end{aligned} \tag{5.17}$$

where $a = 10$, $b = 28$, and $c = 8/3$. As in Sec. 5.3.1, the time series $x(t)$, $y(t)$ and $z(t)$ are available during the training phase, $0 < t < T$. Also, as in Sec. 5.3.1, time series are pre-processed as in Eq. (5.14) to give $\tilde{x}(t), \tilde{y}(t)$ and $\tilde{z}(t)$.

First we consider the case where $\mathbf{u}(t) = \tilde{x}(t)$. After the training phase is over, the reservoir can only access the series $\tilde{x}(t)$ and is able to infer $\tilde{y}(t)$ and $\tilde{z}(t)$ accurately as shown in Figs. 5.5. Clearly, there is no observability problem for $\mathbf{u}(t) = \tilde{x}(t)$.

However, if $\mathbf{u}(t) = \tilde{z}(t)$, then observability does not apply, i.e., it is not possible, even in principle, to infer $\tilde{x}(t)$ from $\tilde{z}(t)$. To see this, we note that the Lorenz system is invariant under the transformation $(x, y, z) \to (-x, -y, z)$. Because of this symmetry, any given time series measurement $z_0(t)$ can correspond to either one of two possible state time series on the attractor $(x_0(t), y_0(t), z_0(t))$ and $(-x_0(t), -y_0(t), z_0(t))$. This results in the system's 'non-observability' when only the z coordinate is measured. Figure 5.6(a) shows that with $\tilde{z}(t)$ as input the reservoir output for $\tilde{x}(t)$ (solid line) completely fails to follow the actual $\tilde{x}(t)$ trajectory (dashed line). A similar result applies for $\tilde{y}(t)$.

However, since the ambiguity in the state lies purely in the sign of the x and

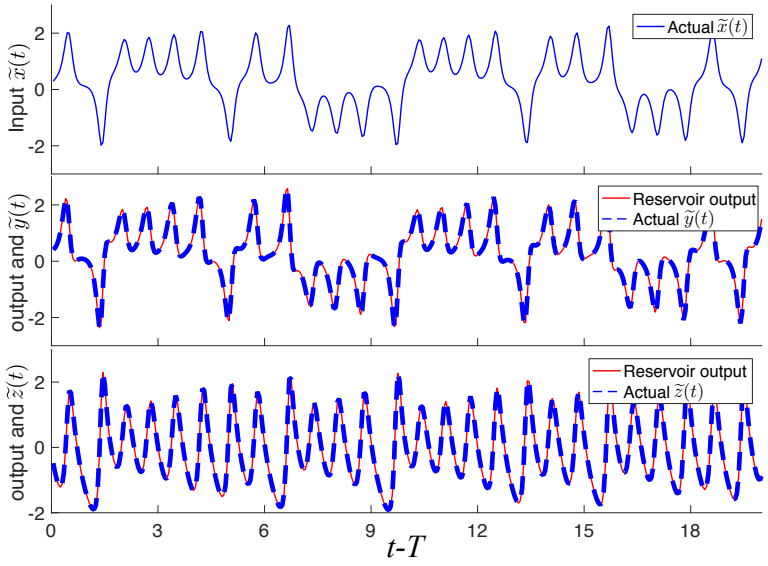


Figure 5.5: For the Lorenz system Eqs. (5.17) with input $\tilde{x}(t)$, the reservoir computer very accurately generates $\tilde{y}(t)$ and $\tilde{z}(t)$. The reservoir parameters we use here and in Figs. 5.6 and 5.7 are the same as those given in Eqs. (5.15) except that $\Delta t = 0.05$.

y variables, it should be possible, in principle, to infer the state variables (x^2, y^2) unambiguously given measurements of the z variable. We show in Fig. 5.6(b) that this is indeed the case.

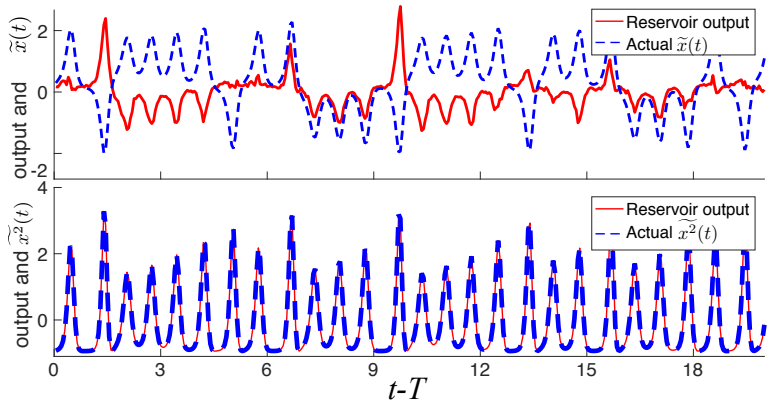


Figure 5.6: (a) A reservoir computer with input $\tilde{z}(t)$ from the Lorenz system Eqs. (5.17) can not generate $\tilde{x}(t)$. (b) However, it can be trained to accurately generate $\widetilde{x^2}(t) = [x^2(t) - \langle x^2(t)\rangle]/\langle[x^2(t) - \langle x^2(t)\rangle]^2\rangle^{1/2}$, and, as in Eq. (5.14), the angle brackets denote time averages.

Another way to demonstrate the interplay, in this example, between the sym-

metry of the system and observability is to modify the Lorenz system slightly in a way that breaks the symmetry. The modified set of Lorenz equations are given by Eqs. (5.17), but with an extra term $x(t)$ added to the right hand side of the third equation (i.e., the equation for dz/dt). With this change a visualization of the attractor (not shown) in (x, y, z) state space shows that the attractor is slightly perturbed from the picture obtained for the original Lorenz system, now with the size and orientations of the two lobes (wings of Lorenz's 'butterfly attractor') being slightly different from each other. As shown in Figs. 5.7, the reservoir is now able to infer the x and y state variables from measurement of the z variable.

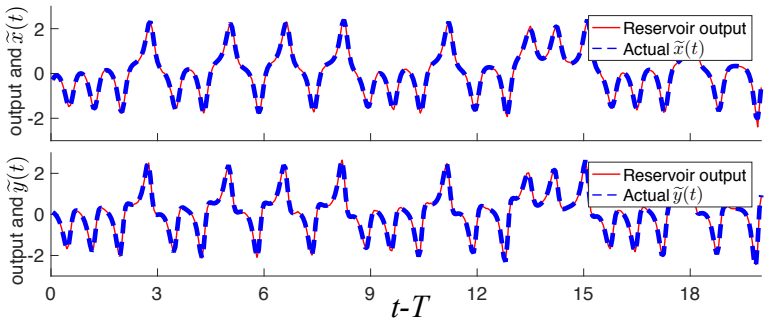


Figure 5.7: A reservoir computer with input $\tilde{z}(t)$ from the modified Lorenz system with its symmetry broken. In this case, the reservoir accurately generates $\tilde{x}(t)$ and $\tilde{y}(t)$.

It is interesting to note that, while, as shown in Fig. 5.5, inference of z when either x or y is given can work well, inference of z completely fails when $\xi = 0$. The reason for this is interesting. Since the hyperbolic tangent in Eq. (5.1) is an odd function, when $\xi = 0$, the composite system consisting of reservoir plus the Lorenz equations has the overall symmetry $(x, y, z, \mathbf{r}) \to (-x, -y, z, -\mathbf{r})$. Thus, we see that, although (x, y) and $(-x, -y)$ may correspond to the same value of z, this is incompatible with the linear output dependence hypothesized in Eq. (5.2), since

$\mathbf{r} \to -\mathbf{r}$ would necessarily change the output estimate of z. However, if $\xi \neq 0$ the symmetry of the composite Lorenz-reservoir system is broken, and inference of z becomes potentially possible. (Even with $\xi = 0$, another alternative, which we have found to work, is to introduce terms quadratic in $\mathbf{r}$ into Eq. (5.2).)

We remark that the symmetry mechanism we focused on in this subsection is meant only as one particular example of the very many ways that nonobservability can arise.

5.4 Conclusions

In this chapter we investigate the application of reservoir computing to infer unmeasured state variables of a chaotic dynamical system from a limited set of continually measured state variables for situations in which a mathematical model of the dynamical system is unavailable or is insufficiently accurate. Our main results are as follows.

(1) Extremely accurate results can be robustly obtained over a wide range of reservoir parameters as shown in Sec. 5.3.1 for the Rössler system.

(2) The issue of 'observability' and the limitation it implies for achieving our goal was investigated in an example (Sec. 5.3.2).

Bibliography

[1] Wikipedia, *Emergence — wikipedia, the free encyclopedia* (2017), [Online; accessed 25-May-2017], https://en.wikipedia.org/w/index.php?title=Emergence&oldid=778942549.

[2] A. Neishtadt, in *Akademiia Nauk SSSR Doklady* (1975), vol. 221, pp. 301–304.

[3] A. Neishtadt, Journal of Applied Mathematics and Mechanics **39**(4), 594 (1975), http://www.sciencedirect.com/science/article/pii/002189287590060X.

[4] A. Neishtadt, Soviet Journal of Plasma Physics **12**, 568 (1986).

[5] J. L. Tennyson, J. R. Cary, and D. F. Escande, Phys. Rev. Lett. **56**, 2117 (May 1986), http://link.aps.org/doi/10.1103/PhysRevLett.56.2117.

[6] J. R. Cary, D. F. Escande, and J. L. Tennyson, Phys. Rev. A **34**, 4256 (Nov 1986), http://link.aps.org/doi/10.1103/PhysRevA.34.4256.

[7] J. Hannay, Journal of Physics A: Mathematical and General **19**(17), L1067 (1986).

[8] A. N. Vasil'ev and M. A. Guzev, Theoretical and Mathematical Physics **68**(3), 907 (1986).

[9] B. V. Chirikov, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **413**(1844), 145 (1987).

[10] J. Henrard, in *Dynamics Reported* (Springer Berlin Heidelberg, 1993), vol. 2 of *Dynamics Reported*, pp. 117–235, http://dx.doi.org/10.1007/978-3-642-61232-9_4.

[11] B. Chirikov and V. Vecheslavov, Journal of Experimental and Theoretical Physics **90**(3), 562 (2000), http://dx.doi.org/10.1134/1.559139.

[12] Y. Elskens and D. Escande, Physica D: Nonlinear Phenomena **62**, 66 (1993), http://www.sciencedirect.com/science/article/pii/0167278993902723.

[13] Y. Elskens and D. F. Escande, Nonlinearity **4**(3), 615 (1991), http://stacks.iop.org/0951-7715/4/i=3/a=002.

[14] S. M. Reppert and D. R. Weaver, Nature **418**(6901), 935 (2002).

[15] E. D. Herzog, Nature Reviews Neuroscience **8**(10), 790 (2007).

[16] E. Ott and T. M. Antonsen, Chaos **18**(3), 037113 (2008).

[17] E. Ott and T. M. Antonsen, Chaos **19**(2), 023117 (2009).

[18] T. H. Monk, D. J. Buysse, J. Carrier, and D. J. Kupfer, J. Sleep Res. **9**(2), 101 (2000).

[19] J. M. Beggs and D. Plenz, J. Neurosci. **23**(35), 11167 (2003).

[20] J. M. Beggs and D. Plenz, J. Neurosci. **24**(22), 5216 (2004).

[21] D. Plenz and D. R. Chialvo, arXiv preprint arXiv:0912.5369 (2009).

[22] W. L. Shew, H. Yang, T. Petermann, R. Roy, and D. Plenz, J. Neurosci. **29**(49), 15595 (2009).

[23] G. Hahn, T. Petermann, M. N. Havenith, S. Yu, W. Singer, D. Plenz, and D. Nikolić, J. Neurophysiol. **104**(6), 3312 (2010).

[24] D. Plenz, The European Physical Journal Special Topics **205**(1), 259 (2012).

[25] S. Yu, H. Yang, O. Shriki, and D. Plenz, Frontiers in Systems Neuroscience **7** (2013).

[26] S.-S. Poil, R. Hardstone, H. D. Mansvelder, and K. Linkenkaer-Hansen, J. Neurosci. **32**(29), 9817 (2012).

[27] N. Friedman, S. Ito, B. A. W. Brinkman, M. Shimono, R. E. L. DeVille, K. A. Dahmen, J. M. Beggs, and T. C. Butler, Phys. Rev. Lett. **108**, 208102 (May 2012), http://link.aps.org/doi/10.1103/PhysRevLett.108.208102.

[28] M. Lukoševičius and H. Jaeger, Computer Science Review **3**(3), 127 (2009).

[29] S. Vaikuntanathan and C. Jarzynski, Phys. Rev. E **83**, 061120 (Jun 2011), http://link.aps.org/doi/10.1103/PhysRevE.83.061120.

[30] R. Marathe and J. M. R. Parrondo, Phys. Rev. Lett. **104**, 245704 (Jun 2010), http://link.aps.org/doi/10.1103/PhysRevLett.104.245704.

[31] J. Parrondo and L. Granger, *Maxwell demons in phase space*, 1504.05388 [cond-mat.stat-mech].

[32] C. R. Menyuk, Physical Review A **31**(5), 3282 (1985).

[33] J. R. Cary and R. T. Skodje, Physica D: Nonlinear Phenomena **36**(3), 287 (1989).

[34] E. Forest and R. D. Ruth, Physica D: Nonlinear Phenomena **43**(1), 105 (1990).

[35] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, 2002), chap. 6, 2nd ed.

[36] E. Ott, Phys. Rev. Lett. **42**, 1628 (Jun 1979), `http://link.aps.org/doi/10.1103/PhysRevLett.42.1628`.

[37] C. Jarzynski, Phys. Rev. Lett. **71**, 839 (Aug 1993), `http://link.aps.org/doi/10.1103/PhysRevLett.71.839`.

[38] M. Berry and J. Robbins, Proc. R. Soc. A **442**(1916), 659 (1993).

[39] S. Yamaguchi, H. Isejima, T. Matsuo, R. Okura, K. Yagita, M. Kobayashi, and H. Okamura, Science **302**(5649), 1408 (2003).

[40] D. K. Welsh, D. E. Logothetis, M. Meister, and S. M. Reppert, Neuron **14**(4), 697 (1995).

[41] S. Honma, W. Nakamura, T. Shirakawa, and K.-i. Honma, Neurosci. Lett. **358**(3), 173 (2004).

[42] S. H. Strogatz, J. Math. Bio. **25**(3), 327 (1987).

[43] A. Gundel and M. B. Spencer, Chronobiology International **9**(2), 148 (1992).

[44] A. Gundel and M. B. Spencer, J. Bio. Rhythms **14**(6), 517 (1999).

[45] Y. Yamaguchi, T. Suzuki, Y. Mizoro, H. Kori, K. Okada, Y. Chen, J.-M. Fustin, F. Yamazaki, N. Mizuguchi, J. Zhang, *et al.*, Science **342**(6154), 85 (2013).

[46] D. Gonze, S. Bernard, C. Waltermann, A. Kramer, and H. Herzel, Biophysical Journal **89**(1), 120 (2005).

[47] S. Bernard, D. Gonze, B. Čajavec, H. Herzel, and A. Kramer, PLoS Comput. Biol. **3**(4), e68 (2007).

[48] M. Hafner, H. Koeppl, and D. Gonze, PLoS Comput. Biol. **8**(3), e1002419 (2012).

[49] J. Myung, S. Hong, D. DeWoskin, E. De Schutter, D. B. Forger, and T. Takumi, Proceedings of the National Academy of Sciences **112**(29), E3920 (2015).

[50] D. DeWoskin, J. Myung, M. D. Belle, H. D. Piggins, T. Takumi, and D. B. Forger, Proceedings of the National Academy of Sciences **112**(29), E3911 (2015).

[51] D. B. Forger and C. S. Peskin, Proceedings of the National Academy of Sciences of the United States of America **102**(2), 321 (2005).

[52] D. Forger, D. Gonze, D. Virshup, and D. K. Welsh, Journal of biological rhythms **22**(3), 200 (2007).

[53] T. Antonsen Jr, R. Faghih, M. Girvan, E. Ott, and J. Platig, Chaos **18**(3), 037112 (2008).

[54] L. M. Childs and S. H. Strogatz, Chaos **18**(4), 043128 (2008).

[55] R. L. Sack, A. J. Lewy, M. L. Blood, L. D. Keith, and H. Nakagawa, Clinical Endocrinology & Metabolism **75**(1), 127 (1992).

[56] J. Mills, Journal of Physiology **174**(2), 217 (1964).

[57] S. S. Campbell, D. Dawson, and J. Zulley, Sleep **16**(7), 638 (1993).

[58] B. Middleton, J. Arendt, and B. Stone, J. Sleep Res. **5**(2), 69 (1996).

[59] C. A. Czeisler, J. F. Duffy, T. L. Shanahan, E. N. Brown, J. F. Mitchell, D. W. Rimmer, J. M. Ronda, E. J. Silva, J. S. Allan, J. S. Emens, *et al.*, Science **284**(5423), 2177 (1999).

[60] S. A. Brown, F. Fleury-Olela, E. Nagoshi, C. Hauser, C. Juge, C. A. Meier, R. Chicheportiche, J.-M. Dayer, U. Albrecht, and U. Schibler, PLoS Biol **3**(10), e338 (2005).

[61] J. A. Mohawk, C. B. Green, and J. S. Takahashi, Annual review of neuroscience **35**, 445 (2012).

[62] Z. Boulos, S. S. Campbell, A. J. Lewy, M. Terman, D.-J. Dijk, and C. I. Eastman, J. Bio. Rhythms **10**(2), 167 (1995).

[63] J. Aschoff, K. Hoffmann, H. Pohl, and R. Wever, Chronobiologia **2**(1), 23 (1974).

[64] P. J. Mitchell, E. K. Hoese, L. Liu, L. F. Fogg, and C. I. Eastman, J. Bio. Rhythms **12**(1), 5 (1997).

[65] T. L. Shanahan, R. E. Kronauer, J. F. Duffy, G. H. Williams, and C. A. Czeisler, J. Bio. Rhythms **14**(3), 237 (1999).

[66] T. Takahashi, M. Sasaki, H. Itoh, H. Sano, W. Yamadera, M. Ozone, K. Obuchi, H. Nishimura, and N. Matsunaga, Psychiatry and Clinical Neurosciences **53**(2), 257 (1999).

[67] T. Takahashi, M. Sasaki, H. Itoh, W. Yamadera, M. Ozone, K. Obuchi, N. Matsunaga, H. Sano, and K.-I. Hayashida, Psychiatry and Clinical Neurosciences **55**(3), 275 (2001).

[68] R. A. Wever, *The circadian system of man: results of experiments under temporal isolation* (Springer Science & Business Media, 2013).

[69] J. Mills, D. Minors, and J. Waterhouse, J Physiol **285**, 455 (1978).

[70] A. Gundel and H. M. Wegmann, Chronobiol Int **6**(2), 147 (1989).

[71] P. So, T. B. Luke, and E. Barreto, Physica D **267**, 16 (2014).

[72] D. Pazó and E. Montbrió, Phys. Rev. X **4**(1), 011009 (2014).

[73] C. W. Eurich, J. M. Herrmann, and U. A. Ernst, Phys. Rev. E **66**(6), 066137 (2002).

[74] A. Levina, J. M. Herrmann, and T. Geisel, Nature Physics **3**(12), 857 (2007).

[75] A. Levina, J. M. Herrmann, and T. Geisel, Phys. Rev. Lett. **102**(11), 118110 (2009).

[76] D. B. Larremore, M. Y. Carpenter, E. Ott, and J. G. Restrepo, Phys. Rev. E **85**, 066131 (Jun 2012), `http://link.aps.org/doi/10.1103/PhysRevE.85.066131`.

[77] D. B. Larremore, W. L. Shew, E. Ott, F. Sorrentino, and J. G. Restrepo, Phys. Rev. Lett. **112**(13), 138103 (2014).

[78] C. Haldeman and J. M. Beggs, Phys. Rev. Lett. **94**, 058101 (Feb 2005), `http://link.aps.org/doi/10.1103/PhysRevLett.94.058101`.

[79] O. Kinouchi and M. Copelli, Nature Physics **2**(5), 348 (2006).

[80] W. L. Shew and D. Plenz, Neuroscientist **19**(1), 88 (2013).

[81] D. R. Chialvo, Nat. Phys. **6**(10), 744 (2010).

[82] T. Mora, S. Deny, and O. Marre, Phys. Rev. Lett. **114**, 078105 (Feb 2015), `http://link.aps.org/doi/10.1103/PhysRevLett.114.078105`.

[83] M. Nonnenmacher, C. Behrens, P. Berens, M. Bethge, and J. H. Macke, arXiv preprint arXiv:1603.00097 (2016).

[84] J. G. Skellam, Journal of the Royal Statistical Society. Series A (General) **109**(Pt 3), 296 (1945).

[85] R. E. Kalman, Journal of Basic Engineering **82**(1), 35 (1960).

[86] K. Ogata and Y. Yang, *Modern control engineering* (Prentice-Hall, Upper Saddle River, NJ, 1970).

[87] R. Hermann and A. J. Krener, IEEE Transactions on automatic control **22**(5), 728 (1977).

[88] P. So, E. Ott, and W. Dayawansa, Physical Review E **49**(4), 2650 (1994).

[89] J. C. Quinn, P. H. Bryant, D. R. Creveling, S. R. Klein, and H. D. Abarbanel, Physical Review E **80**(1), 016201 (2009).

[90] J. Schumann-Bischoff, S. Luther, and U. Parlitz, Physical Review E **94**(3), 032221 (2016).

[91] O. E. Rössler, Physics Letters A **57**(5), 397 (1976).

[92] E. N. Lorenz, Journal of the atmospheric sciences **20**(2), 130 (1963).

[93] H. Jaeger and H. Haas, Science **304**(5667), 78 (2004).

[94] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert, Neural networks **20**(3), 335 (2007).

[95] D. Brunner, S. Reitzenstein, and I. Fischer, in *Rebooting Computing (ICRC), IEEE International Conference on*, IEEE (2016), pp. 1–2.

[96] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, Nature communications **2**, 468 (2011).

[97] N. D. Haynes, M. C. Soriano, D. P. Rosin, I. Fischer, and D. J. Gauthier, Physical Review E **91**(2), 020801 (2015).

[98] L. Larger, A. Baylón-Fuentes, R. Martinenghi, V. S. Udaltsov, Y. K. Chembo, and M. Jacquot, Phys. Rev. X **7**, 011015 (Feb 2017), `http://link.aps.org/doi/10.1103/PhysRevX.7.011015`.

[99] C. Fernando and S. Sojakka, in *European Conference on Artificial Life*, Springer (2003), pp. 588–597.

[100] F. Takens, in *Dynamical systems and turbulence, Warwick 1980* (Springer, 1981), pp. 366–381.

[101] A. Brandstater and H. L. Swinney, Physical Review A **35**(5), 2207 (1987).

[102] T. Sauer, J. A. Yorke, and M. Casdagli, Journal of statistical Physics **65**(3), 579 (1991).